

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2005, 11:59:50 ; Search time 5661 Seconds
(without alignments)
12809.121 Million cell updates/sec

Title: US-10-757-093-3

Perfect score: 1905
Sequence: 1 atgaattcctcagggatc.....ctagatgataagaatcaa 1905

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_esc1:*
2: gb_esc2:*
3: gb_hic:*
4: gb_esc3:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_gsa1:*
9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	261.2	13.7	921	9	CL468845 SAIL_443
2	222	11.7	2274	3	AK041058 Mus muscu
3	208.6	11.0	566	7	CV169266 remexl_01
4	205.8	10.8	1995	6	CD014092 90134967
5	188.4	9.9	2473	3	AY321342 Rattus no
6	183.2	9.6	1377	3	CR593823 full-length
7	183.2	9.6	1853	6	CD014093 90135266
8	180.4	9.5	857	5	BX745933 BX745933
9	178.8	9.4	811	9	CL679241 PRI0125C
c 10	175.4	9.2	908	9	CL463714 SAIL_119
c 11	166.2	8.7	1124	6	CD503076 CDA60-C07
c 12	164.8	8.7	907	9	CL478429 SAIL_28 D
c 13	158.8	8.3	1138	6	CD503098 CDA60-D07
c 14	155.4	8.2	740	6	CF521612 AGENCOURT
c 15	150.6	7.9	689	7	CR442804 CR442804
c 16	145.6	7.6	914	6	CB203472 AGENCOURT
c 17	141.4	7.4	767	4	CF255373 mevn127_c
c 18	139.2	7.3	906	5	BQ941196 AGENCOURT
c 19	139	7.3	957	5	CL495417 SAIL_60 A
c 20	137.8	7.2	1051	5	BX363460 BX363460
c 21	137.4	7.2	1268	6	CD503077 CDA60-C07
c 22	133.6	7.0	913	9	CL474264 SAIL_215
c 23	131	6.9	857	6	CB559443 AGENCOURT
c 24	130.6	6.9	952	9	CL468844 SAIL_443

25	129.4	6.8	2095	6	CD014094 90135027
c 26	127	6.7	907	9	CL482753 SAIL_367
c 27	126.6	6.6	924	4	BI445636 deeb3e05
c 28	126.4	6.6	963	9	CL461373 SAIL_1146
c 29	126.2	6.6	892	9	CL462844 SAIL_1171
c 30	126	6.6	616	2	BE179118 RCO-HT061
c 31	126	6.6	900	6	CA489317 AGENCOURT
c 32	125.8	6.6	790	7	CO559438 AGENCOURT
c 33	124.8	6.6	667	7	CN264045 170004243
c 34	123.8	6.5	884	9	CL461909 SAIL_1154
c 35	123.2	6.5	802	5	BUS37212 603474094
c 36	122.4	6.4	799	5	BUS36987 603585124
c 37	121.6	6.4	1023	5	BX363459 BX363459
c 38	121.2	6.4	816	7	CK467358 938649 MA
c 39	121	6.4	752	4	BJ733141 BJ733141
c 40	121	6.4	846	6	CA453907 AGENCOURT
c 41	120.6	6.3	791	5	BUS39978 603323761
c 42	120	6.3	603	1	AL676250 AL676250
c 43	119.4	6.3	725	5	BUS15788 BUS15788
c 44	119.4	6.3	777	4	BG741172 BG741172
c 45	119	6.2	639	5	BM997809 UI-H-D10-

ALIGNMENTS

RESULT 1
LOCUS CL468845/c
DEFINITION SAIL_443_B06.v2 SAIL Collection Arabidopsis thaliana linear GSS 01-APR-2004
SAIL_443_B06.v2, genomic survey sequence.

ACCESSION CL468845
VERSION CL468845.1 GI:45965487

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 921)

AUTHORS Sessions,A., Burke,E., Presting,G., Aux,G., McEliver,J., Patton,D., Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D., Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmery,J.B., Mittel,T., Karagiri,F., Glazebrook,J., Law,M. and Goff,S.A.

TITLE A high-throughput Arabidopsis reverse genetics system

JOURNAL Plant Cell 14 (12), 2985-2994 (2002)

MEDLINE 22356987

PUBMED 12468722

COMMENT Contact: Sessions A

Applied Trait Genetics

Syngenta Biotechnology Inc.

3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA

Email: allen.sessions@syngenta.com

ABRC Stock Number CS820387; T-DNA left border flanking sequences of

Syngenta Arabidopsis insertion library (SAIL) lines are available

through the Arabidopsis Biological Resource Center (ABRC).

Sequences represent a pool of amplified genomic regions and not

single contiguous sequences.

Class: TDNA tagged.

Location/Qualifiers

1..921

source

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/ecotype="Columbia"

/db_xref="taxon:3702"

/clone="SAIL_443_B06.v2"

/clone_lib="SAIL Collection"

/note="T-DNA left border sequences were isolated using a

modified TAIL-PCR strategy"

ORIGIN

Query Match

Best Local Similarity

13.7%; Score 261.2; DB 9; Length 921;

61.2%; Pred. No. 1.6e-70;

Matches	442;	Conservative	0;	Mismatches	273;	Indels	7;	Gaps	1;
QY	1160	GAATTCGTCGATCGATGAAACACCTGCCGTGATCGAATTCGCTTGATGGCGTAT	1219						
Db	795	GCATCTGGTGGATGATGAAACCTGCTCTGCTGCTTAACTCTCTTGAGCATTTGGT	736						
QY	1220	CTGAGAGTGTGCC-----CCACAACATTACGCCAGATSCGATTAACGATAAAC	1272						
Db	735	TCGAGCGGGGCAACAGCCGAAAGAACTGTACAGGAAAGGACAGTCAACGGGGAAC	676						
QY	1273	CNAGAGCCCAACAGCAGGAGATTCGTGAGCTCATTTGCCGACACAAACCATG	1332						
Db	675	CACCAAGCCGACCTTACAGGGGATTAAAGACTATAGCGGTACAAACCCCAAGC	616						
QY	1333	GTTGTCATGTGCTCTATTGCGCAAGCCCGCATCTCATAGATGAGTCCCGAATAC	1392						
Db	615	GTGGTATGTGGAGTATTTGCCAACGAAACCGGATACCGGTCCGAGGTGACGGGA	556						
QY	1393	TTGAGCCCACTGACCAATTGACTCGTCAACTTGATCAACTGCGCTTATTAAT	1452						
Db	555	TTGCGCGCACTGGCGGAAGACGCGTAACTGACCCGACGCGTCGATCACTGG	496						
QY	1453	AAAGTGGGACGCGGACATTCAGCTGATCGGATCTCTGATCTGTTGATGTCAG	1512						
Db	495	AATGTAATGTTCTGCAACGCTCACCGATACATGACGATCTCTTGATGTCGTGC	436						
QY	1513	ATAAATCGTATTTGCGATGATGATCTCAACAGAGACCTTGAGGAGCAGGAGCT	1572						
Db	435	CTGAACCGTTTATTCAGATGATGATGTCGAAAGCGCGATTTGAAACGCGAAG	376						
QY	1573	CTTGAAGAGAGTGCATGATGATGATGATGATGATGATGATGATGATGATGAT	1632						
Db	375	CTGGAAGAGAGATCTGCGCTGCGGAGAGAACTGATGATGATGATGATGATGAT	316						
QY	1633	TATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1692						
Db	315	TATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	256						
QY	1693	TTTCAAGTACAAATGCTAGATGATGATGATGATGATGATGATGATGATGAT	1752						
Db	255	TATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	196						
QY	1753	GCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1812						
Db	195	GCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	136						
QY	1813	GCTAAGAGAGAGTGTGTTTACCGCTGACCGGAAAGCGGAGCTCATGTTG	1872						
Db	135	GCTAAGAGAGAGTGTGTTTACCGCTGACCGGAAAGCGGAGCTCATGTTG	76						
QY	1873	AG 1874							
Db	75	AG 74							

RESULT 2
AK041058
LOCUS
DEFINITION
AK041058 2274 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male aorta and vein cDNA, RIKEN full-length
structural library, clone:AS30072005 product:beta-glucuronidase
structural, full insert sequence.
ACCESSION
AK041058
VERSION
AK041058.1 GI:26088231
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
MEDLINE
99279253

PUBMED
REFERENCE
AUTHORS
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
MEDLINE
20499374
PUBMED
11042159
REFERENCE
AUTHORS
3
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Komno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL
MEDLINE
20530913
PUBMED
11076861
REFERENCE
AUTHORS
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
JOURNAL
MEDLINE
20530913
PUBMED
11076861
REFERENCE
AUTHORS
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
JOURNAL
MEDLINE
20530913
PUBMED
11076861
REFERENCE
AUTHORS
6
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, I., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct SubMISSION
Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
LOCATION/Qualifiers
1. 2274
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM:DB:A530072005"
/db_xref="taxon:10090"
/clone="A530072005"
/sex="male"
/tissue_type="aorta and vein"
/clone_lib="RIKEN full-length enriched mouse cDNA library"

```

misc_feature      /dev_stage="adult"
                  14..1778
                  /note="beta-glucuronidase structural [MGDI:95874,
                  GB|NM_010368, evidence: BLASTN, 99%, match=733]
                  putative"
polyA_signal      2260..2265
                  /note="putative"
polyA_site        2274
                  /note="putative"

ORIGIN
Query Match      11.7%; Score 222; DB 3; Length 2274;
Best Local Similarity 51.3%; Pred. No. 6e-58;
Matches 695; Conservative 0; Mismatches 590; Indels 69; Gaps 5;

QY 563 GGAACCGCATGGCAAGAAATCCACACTATCAATGACTTTTCAACTATGCTGTG 622
DB 409 GGTATCCCAAGGGTTATCTTGTCCAGACACAGCTTTGACTTTCATATGCGGAGC 468
QY 623 TCGCCGATCTATCTGGCTTTATTTCTGTACCCAGCAACATATCCAGATATTACTGTG 682
DB 469 TGCATGATCTGTGTCTCTATATCAACCCCTTACCTTACATCATGATATATCATCTGTGA 528
QY 683 TTACAGATGTGAATGTGCAATGTGTCTGATTATACAGAGGTGCAAGTGGCGAACAGA 742
DB 529 TCACATAATGTGAGCAAGACATCGGGCTGGTACCTAGATTTCTGTGACAGGCGAGTG 588
QY 743 CGACGGGGCAATCCAGATCTCAGTATGACAGAGATGAGAGTATTTGTTCAGAGCCCT 802
DB 589 AACATTTCCAGCTAGAAAGTGCAACTTTTGATGAGGTGGCAAAAGTCGGGCCCATGGGA 648
QY 803 CGGAGACTCAGGGTACTGTCACAATTCCTCAGTCAAGCTATGGCAACCTGGCGCCGAT 862
DB 649 CAGGGAACCAAGGCTCAACTTCAGGTTCCAGTCCAACTCTGTGGCTTACCTGATGC 708
QY 863 ATCTCTACCACTCCAGGTCAAC-----ATCGTGGTTCTAGCGGAGATGTAG 910
DB 709 ATGAGCATCCACGCTCATGTACTCTGTGAGGTGAAGTGACAAACAATGATCTGTGA 768
QY 911 TCGACACCTTCAATTTGGCTACGGGGGTGCTACTGTCAAGTGTCCGGGTCACAATTC 970
DB 769 CTGACTACTTCAACCCCTCTATCGGGATTCGAACATGTGCTGTCAAAAGCAACATTC 828
QY 971 TATATAATGAAAGCCTTTCTACTTTACCGGTTTGGCAAAACATGAAGCACAGCATGAC 1030
DB 829 TCAATAACGGGAAGCCCTTCTATTTCCAAAGGGGTCAATAGACAGAGATTCAATATCC 888
QY 1031 GTGGCAAGAGACATGACCCAGACATACATGTTCACAGATTTCCAACTCATGAATGATTTG 1090
DB 889 GAGGGAAGAGCTTCGACCTGGCGGTCTGTGAAGATTTCAACGTGCTCCGTGCTCG 948
QY 1091 GAGCAAAATCTTTTGGCACTTCACTATTCCTTACCGGGAAGGTTCATGATTTGCGAG 1150
DB 949 GGGCAAAATTCCTTTCGTACCAAGCACTATTCCTTCACTCAAGAGGTACTTCAAGCTCTGTG 1008
QY 1151 ATCGAAATGGAATGTGCGATGATGAAGAAACCTGCGGTGCTGAACATTGCTTGA 1210
DB 1009 ACCGATACGGAGTTGTGTATCATGATGAGTGTCCCGGTGGG-----CA 1053
QY 1211 TGGGCGTATCTGAGATGTGCCCCCAAAACATTTTACGCCAGATGCGATTAAAGATMAAA 1270
DB 1054 TTGTGTACTCTAGAGTTTGGCAAGAGTCACTTGGC-----1092
QY 1271 CCCAAGAGGCCCAAGACAGCGGATTCGTGAGTCACTTCCCGAGACAAAACCATCCA 1330
DB 1093 -----ACCACTTAGAGGTATGAGAGAGCTGTGCTCCCGGACAAAACATCACTCTG 1143
QY 1331 GTGTTGTCATGTGTCTATTTGCAAGAGCCGATCTCATGAAGATGAGTCTCGGAAT 1390
DB 1144 CGGTTGTATGTGTCTGTGGCCAAATGAGCTTCTCTGTCTGAAAACCGCGCATATT 1203
QY 1391 ACTTCGAGCACTGACCAATTTGACTCGTCAACTTGAATCAACTCGCCCTATTATTAATTTG 1450

```

```

DB 1204 ACTTAAAGCGCTGATCAACCCACCAAAAGCCCTGAGCCTCAACCCGTCCGTGACCTTT- 1262
QY 1451 CTAACTCGGACAGGGGACATATACAGCTGAGATCGATCTCTGATCTGTTGATGTGAGTT 1510
DB 1263 -----GTAGCAACGCCAATATGATGACAGCTGGGGGCCCCGTACGTGATGTATCT 1317
QY 1511 GCATAAATCGGATATTTCCGATGTATTTCTCAACAGAGAACCTTGAGAAAGCAGGCGAG 1570
DB 1318 GTGTAAACAGCTACTTTCTGTGTATCATGACTATGGGCAATTTGAGGTGATTCAGCCAC 1377
QY 1571 CTCTGAAAAGAGCTGACATGATGCGCAAGAGAAATTCACAGCCGATCTCATACCG 1630
DB 1378 AGTGAATGCGCAGTTTGAACCTGTATTAAGCCGATCAGAGCCGATTTATCCAGAGG 1437
QY 1631 AATATGTCAGATACCTTGGAGGCTTCACTCTATCTCCGAGCTCCGTTGAGCGAAG 1690
DB 1438 AGTATGACAGAGCGCAATCCAGGATCAGAGAGACCCGCTGCATGTTCAATGAGG 1497
QY 1691 AGTTCCAACTACAAATGTGACATGTATCACTGATGATTTGAT-----CGCATTTAGT 1744
DB 1498 AGTACAGAAAGGCTGTTCGAGAAATTAACATTCAGTTCTGTGATCAAGAAAGTAAAT 1557
QY 1745 CGATGCGAGCGACATGTTTGAATCTGCCGATTTCAAGACCAACTTGGGTATCATCC 1804
DB 1558 ACGTGTCCGAGAGCTCATCTGGAATTTCCGACATTCAGTACGAAACGATCACCGCTGA 1617
QY 1805 GAGTGAACGTTAACAAGAGGGTGTTCACCCGTGACCGAAAGCCAAAGGCGGCGACTG 1864
DB 1618 GAGTAAATCGAAACAAAGAGGGATCTTCACTGCGCAAGACAGCCCAAACTTCGGCCT 1677
QY 1865 ATAGTTGAGGCGCAAGGTGACTAGTATTGATTA 1898
DB 1678 TTATTTGGAGAGATATCTGAGGATTTGCCAA 1711

```

```

RESULT 3
CVI69266      566 bp  mRNA  linear  EST 09-SEP-2004
LOCUS      rsmxl1.010419.Y1.rcf cDNA library of Salvia miltiorrhiza Salvia
DEFINITION  miltiorrhiza cDNA 5', mRNA sequence.
ACCESSION  CVI69266
VERSION    CVI69266.1 GI:51958406
KEYWORDS   EST.
SOURCE     Salvia miltiorrhiza
ORGANISM   Salvia miltiorrhiza

REFERENCE
1 (bases 1 to 566)
Wang, Z., Yan, Y. and Tian, W.
Gene Discovery in Salvia miltiorrhiza Bge. as Revealed by EST
Sequencing
Unpublished (2004)
CONTACT: Zhezhong Wang, Yaping Yan, Wei Tian
Plant Cell Engineering
(Tian) Hangzhou Genomics Research and Development Institute
No. 199 of Chang'an South Road, Xi'an, Shaanxi, 710062, The
People's Republic of China
Tel: 86-29-85308352
Fax: 86-29-85303736
Email: zzwang@smnu.edu.cn
PCR Primers
FORWARD: M13 R
BACKWARD: M13 F
Seq primer: 3'-gtacagatcgacaa-5' reverse primer M13.
Location/Qualifiers
1..566
/organism="Salvia miltiorrhiza"
/mol_type="mRNA"
/db_xref="taxon:226208"
/sex="hermaphrodite"

```

```

FEATURES
source
1..566
/organism="Salvia miltiorrhiza"
/mol_type="mRNA"
/db_xref="taxon:226208"
/sex="hermaphrodite"

```

/tissue type="Whole Youth Plant"
 /dev stage="cultured about 1mon-3mon"
 /clone lib="CDNA Library of Salvia miltiorrhiza"
 /note="Organ: Whole Youth Plant; Vector: pBlueScript;
 Site 1: EORI; Site 2: XhoII; mRNA isolation used
 PolyAtract (r) mRNA Isolation Systems (PROMEGA, CAT
 No.25200) and pBlueScript (r) XR CDNA Library Construction
 Kit (Invitrogen, CAT No200455.) to construct the cDNA
 library."

ORIGIN

Query Match 11.0%; Score 208.6; DB 7; Length 566;
 Best Local Similarity 62.6%; Pred. No. 6.4e-54;
 Matches 325; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 1251 AGATGCGATTAAACATTAACCCAGAGGCCCAACAGCCGATTCGTGAGCTCATTCG 1310
 DB 48 AGAGGCGATGCAACGGGAACTCGACAGCGCACTTACAGCCATTAAAGAGCTGATAGC 107
 QY 1311 CCGAGCAAAAACCCAGCGGTGTCATGTCATTTGTCACCAAGCCCGCATCTCA 1370
 DB 108 GCGTGACAAAACCCAGCGGTGTCATGTCATTTGTCACCAAGCCCGCATCTCA 167
 QY 1371 TGAAGATGAGCTCGGCAATTAATTCGAGCCATGACCAATTTGACTGTCACTTGATCC 1430
 DB 168 TCGGCAAGGTGACGGGAATATTTCCGCCACTGGCCGAGCAACCGCTAACTCGACCC 227
 QY 1431 AACTCGCCCTATTACATTTCTTAACGTCCGCAAGCGACATCATGAGTCGATCTC 1490
 DB 228 GAGCGCTCCATACCGCTGCTCATATGTTCTGAGACGCTCAACCGATACATCAG 287
 QY 1491 TGATCTGTTGATGTCATGTTCAATTAATTCGATGATGTTCTTCAACAGAGA 1550
 DB 288 CGATCTCTTTGATGTCGTGCTGCGTGAACCGTTATTCAGATGATGTTCAACAGCGCA 347
 QY 1551 CCTTGAAGAGCAGAGGCGACTCTTGAAGAGCGTCGACATGATGCAAGAAATTCCA 1610
 DB 348 TTGGAAGAGCGCAGAGAGGTACTGGAAGAAATCTTGCGCTGCGAGAGAACTGCA 407
 QY 1611 CAGGCGCATGTCATGACCGAATATGTCAGATACCCCTTGACGCTTCACTTACCT 1670
 DB 408 TCAGCGCATATCATCACCGAATACGCGGTGATAGCCGGCTGACCTCATATGA 467
 QY 1671 CGAGATGCTTGAAGCAAGATTTCCAGTACAAATGCTAGACATGATACCATGATGTT 1730
 DB 468 CACCGCATGTGAGTGAAGAGATATCATGTCATGCTGATATGATACCGCGTCTT 527
 QY 1731 TGATCGCATGATGATGTCAGGCGAGCATGTTTGA 1769
 DB 528 TGATCGCGTACGCGCTGCTGCTGTAACAGTATGGA 566

RESULT 4
 LOCUS CD014092 1995 bp mRNA linear EST 21-OCT-2003
 DEFINITION 90134967 Single gene library Homo sapiens cDNA, mRNA sequence.
 ACCESSION CD014092
 VERSION CD014092.1 GI:3777621
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1995)
 Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,
 Au-Young, J. and Stuve, L.L.
 PCR isolation and cloning of novel splice variant mRNAs from known
 drug target genes
 Genomics 83 (4), 566-571 (2004)
 COMMENT
 JOURNAL
 CONTACT: Jin, P.
 Incyte Corporation
 3160 Porter Drive,
 Palo Alto, CA 94304, USA
 Tel: 650 621 8639

Fax: 650 621 8965
 Email: pjin@incyte.com.
 Location/Qualifiers
 FEATURES
 source

1.1995
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone lib="Single gene library"
 /note="Vector: pDrive Cloning Vector; RT-PCR was performed
 using gene-specific primers flanking the open-reading
 frame. PCR products were subcloned into pDrive Cloning
 Vector and sequenced completely using M13 forward and
 reverse primers. Sequencing gaps were closed by
 re-sequencing using primers flanking the gapped areas."

ORIGIN

Query Match 10.8%; Score 205.8; DB 6; Length 1995;
 Best Local Similarity 50.1%; Pred. No. 7.7e-53;
 Matches 688; Conservative 0; Mismatches 607; Indels 78; Gaps 4;

QY 553 ATCAGCAGAGGAGCGCGCTGCAAGAGATTCAGACTATCAATGACTTTAACA 612
 DB 399 ATGCCATGCTGATATCCAGAGGTTACTTGTCCAGAACATATTTGACTTTTCAAC 458
 QY 613 TATGCTGATCTGCGCCGATCTATCTGCTTATTTCTGTACCCAGCAACATATCCAGAT 672
 DB 459 TAGCTGACCTCAGAGGCTGTACTTCTGTACAGACACCCACACCTACATCATGAC 518
 QY 673 ATTACTGTGTGTTACAGATGTTGATGTCATATGCTGTGATTAATCTAGAGGTCAAGTG 732
 DB 519 ATCACCGTACACCAAGCGGTGAGCAAGACATGTCGGCTGTGTAATTAACAGATCTCTGTC 578
 QY 733 GCGAACACAGACAGCGGGGAGATCCAGATCTCAGATGTCAGAGATGAGACTATGTT 792
 DB 579 AAGGCGATTAACCTGTTCAAGTTGGAAGTGCCTTTTGATGCGAAGAAACAAAGTCGT 638
 QY 793 GCAAGGCTCTGAGGCTCAGGTACTGTCAATTCCTCAGTCAAGCTATG----- 846
 DB 639 GCGAATGGGACTGGGACCCAGGGCCAATTAAGTGCCAGGATGACGCTCTGGTGGCG 698
 QY 847 -----CAACCTGGGCGCCATATCTCTCAACTCCACTCAGTCAACATCTGT 891
 DB 699 TACTGATGACGAAGCGCTGCTCTATCTGTATTCATTTGAGAGTGCACCTGACACAG 758
 QY 892 GGTTCAGGCGGAGATGATGTCACACTTACAAATTTGGCTACGGGGCTCGTATGTCAG 951
 DB 759 AGTCACTGGGCGCTGTGTGATCTTACACACTCCCTGTGGGATCCGACATGTGCT 818
 QY 952 GTTCCGGGATCAATTTCTTAATAAGAAAGCCTTTCTACTTAAACGGTTTGGCAAA 1011
 DB 819 GTCAACCAAGAGCAGTTCCTCATCAATGGGAAACCTTTCTATTTCCACGGTGTCAAG 878
 QY 1012 CATGAAGACACAGAGTACGTGCGAAGAGACATGACCCAGATACATGTTCAAGATTTTC 1071
 DB 879 CATGAGATGCGGACATCCGAGGAGGCTTTCAGTCGGCGCTGCTGTGAAGAACTTC 938
 QY 1072 CAACTCATGAATGATGATGAGCAAAATCTTTTGGCACTTCACTATCTTTAGCGGGA 1131
 DB 939 AACCTGCTCGCTGCTGTGTGTCACACCTTTCCGTACAGCCACTACCTCTATGACAG 998
 QY 1132 GAGGTGATGATTTGCGAGATGCAATGGAATGATGTCATGATGAAACCTGCGCTT 1191
 DB 999 GAAATGATGACAGATGTGACCGCTATGAGATGTGTGTCATGATGATGATGCCGGCTG 1058
 QY 1192 GGTGGAACATTTGCTTGAATGGGGTATCTGAAGTGTGTCACCAAACTTTACGCCA 1251
 DB 1059 GGCCTGGCGCT-----GCCG 1073
 QY 1252 GATGCGATTAACGATTAACCCAGAGGCCACAGAGCGGATTCGTGAGCTCATTTGCC 1311
 DB 1074 CAGTTCTTCAACAACGTTTCTCTGATCACAACATGACAGATGAGAAAGTGTGCT 1133
 QY 1312 CGAGCAAAAACATGCCAGTGTGTCATGATGTCATATGCGAAGAGCCGCGATCTCAT 1371

Db	1134	AGGACAAAGAACCA	CCCCCGCGTGGTAGTGATCTTGTCGCCAAAGACCTGCCTCCAC	1199
Oy	1372	GAAAGTAGAGCTCGGAA	TACTTCAGGCCA CTGACCAATTGAC TCGCAACTTGANTCA	1431
Db	1194	CTAGAATCTGCTGGCTA	CTACTTGAAAGATGGTATCGCTCACA CAAATCCTTGGACC CC	1255
Oy	1432	ACTGCGCCCTATTA	CACTTTGGTCTACGTCCGCA CGCGACATCATCAGCTGGATCTCT	1491
Db	1254	TCCCGGCTCTGACCTTT	-----GTGAGCAA CTCCTAATGACAGACA CAAGGAGGCGCT	1307
Oy	1492	GATCGTTGATGTGATG	CATAATGCGTATTTCCGATGGTATTTCA AAGAGAGAC	1551
Db	1308	CCGATATGTGAATGTAT	CTTTTGAACAGCTACTCTCTTGATACGACTACGACGACAC	1367
Oy	1612	AGCGCGATCGTCAT	ATACCGGAATATGTGTGACATATACCTTCGACGCGCTTCACTTACCTC	1671
Db	1428	AAGCCCATTAATYCA	GAGGAGATGTGACAGAAAGATTTGACAGGGTTTCCACAGAGATCCA	1487
Oy	1672	GGACTGCTTGAGAG	GAGATGTCCAAATGATCTAGACATGTACATCGAGTGTTT	1731
Db	1488	CCTGTGATGTTCA	CTGAAGATGACCAAAAAGCTCTGTAGACATACATCTGGGCTCTG	1547
Oy	1732	GAT-----CGCAT	TGAGTCGATGCGACAGCGAGCATGTTTGGAACTTCGCGGATTTCCAG	1785
Db	1548	GATCAAAAACGAGA	AAATATAGTGTGATGAGAGCTCACTTTGGAAATTTGCCGATTTTCAATG	1607
Oy	1786	ACCAACTTGGGTAT	CATCCGATAGACGCTTA CAAAGGCTGTTTTCA CCGCTGACCGCA	1845
Db	1608	ACTGAACAGTCA	CCGACGAGTGTGGGGGAATTA AAAAGGGGATCTTCACTCGCGCAGAGA	1667
Oy	1846	AAGCAAAAGGGCG	CGACTCATAGTTTGA GGGCAAGCTGAGCTAGATTTGATAA	1898
Db	1668	CAACCAAAAAGTGA	CGGCTTCCTTTTCGAGAGAGATCTGGAAGATTTGCCAA	1720
RESULT 5				
LOCUS	AY121342	2473 bp	mRNA	linear HTC 16-JUL-2003
DEFINITION	Rattus norvegicus Ac2-223 mRNA,	complete cds.		
VERSION	AY121342			
KEYWORDS	AY121342.1 GI:32527744			
SOURCE	HTC.			
ORGANISM	Rattus norvegicus (Norway rat)			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
	Rattus.			
TITLE	1 (bases 1 to 2473)			
JOURNAL	Xu,C.S., Li,W.Q., Li,Y.C., Wang,L., Wang,S.F., Han,H.P., Wang,G.P.,			
REFERENCE	Chai,L.Q., Yuan,J.Y., Yang,K.O., Yan,H.M., Chang,C.F., Zhao,L.F.,			
AUTHORS	Ma,H., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.			
	Liver regeneration after PH			
	Unpublished			
	2 (bases 1 to 2473)			
	Xu,C.S., Li,W.Q., Li,Y.C., Wang,L., Wang,S.F., Han,H.P., Wang,G.P.,			
	Chai,L.Q., Yuan,J.Y., Yang,K.O., Yan,H.M., Chang,C.F., Zhao,L.F.,			
	Ma,H., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.			
	Direct Submision			
	Submitted (11-JUN-2003) Henan Bioengineering Key Lab, Henan Normal			
	University, NO. 148 Jianshe Road, Xinxing City, Henan 453002, P.R.			
	China			
FEATURES				
source	Location/Qualifiers			
	1..2473			
	/organism="Rattus norvegicus"			
	/mol_type="mRNA"			
	/db_xref="taxon:10116"			
CDS	41..2467			
	/note="Liver regeneration related protein LRRG134"			

```

/codon_start=1
/product="Ac2-223"
/protein_id="AAP6274.1"
/db_xref="GI:32527745"
/translation="MRKDFQSGVEKYTLMDLSRVLAQNERDRESEAKRPSHTAGI
PAMSPRSRYCVLGGQLCSALALQGMQLPREPSTRSLQIDGLMSPRADYSNNR
LQEPKRMQYRLREGPTLDMVPSPSSFLIQEATRLNLTGWVYREARVLPPQRM
DTRKRVLRINSAHYIAVAVWANGIHVVEHEGHLPEPADITKLQVSPPLTFFRYTIAI
NNTLTPYTLPPGTTVYKTTDSMYPKGFVODISFDFPNVYAGHSVLAITPTPTIID
IITTTDDRVGLVNWIMSVQSGDHQLEFRLLDEBDCKIARGGNGGOLKVPRAHLM
WPIIMHHPAYILSYLETMTTPSPESDPYTLPGAIRIATVATKFKFLNGRPFVDCG
KHSDSDIRGGGPMWPLLIIDFNLLRMVGLMSPTSHRYVSEEVQLCTDRGIIVDGC
PGCITVLPOSGFNVSLRHLHEVWDELRYLRKKNPAPVYMSVNAEPUSLRCAGYTPCL
ITVQGLKTSDEPTGVALVNAHAPYVDITCVNSYSLAWHDHGLEVTLQQLTSQFEN
WYKMOKPIIQSEYGAADVASGLHEDPRMSEBQYALHENVHLIDBEKREYVIGELI
INAFDAFMNTQISRLVYGNKKGIETROBNQKMAPIALIREBMYIINMETRLQIYNQSF
VVDDEKEVQHICGADBDALQLQHKVQNVGDSDVIAGLEICSYCTNRRAHELCIKST
VPSQDQAGORGGQLSIAIGTGMVAVTIKACSHHQVQLKTTVAHSLIAYVKPDRPLRS
CLLDISAGSA"

```

ORIGIN

Query Match	9.9%;	Score 188.4;	DB 3;	Length 2473;
Best Local Similarity	49.0%;	Pred. No. 2.7e-47;		
Matches 833; Conservative	0;	Mismatches 756;	Indels 111;	Gaps 8;

Qy	247	CCAGTCCCGGCGCTTTTCAACGACATCTTATCATGCGGGAGATTCAAGCATGTGGGA	306
Db	413	CCGGTCCCTTCAGAGCTTCATATGACATACCCAGAAGACAGGCTTGGAACTTCATTGGC	472
Qy	307	TGGGTTACTATCATAGCTGAGGTGATCTGTCCCAAGGCGTGCTCAGGA-----G	357
Db	473	TGGGTGTGTATGAAACGGGAAGACAGTGCCTTCCACAGGAATGAGCCCGACACCGACGG	532
Qy	358	CGATATCTCGTGGAGCGCAATCCGCTACGCACCATGATGGATGCATCTATGTCAACACGG	417
Db	533	AGAGTGTGTGTGAGAAATCAACAGCGCCATTACTATGCACTTGTGTGGTGAATGGGATT	592
Qy	418	CTTGTTGCCGAGCATGTGGCGGCGCTATATACCTTTTGAAACGGAAGCTCACTGAATTATC	477
Db	593	CATGTGTGTGAACATGAGGAGGTTCACCTCCCTTTGAGCTGACATCACCAAGCTGTGC	652
Qy	478	-----GCCCCGGAGAGAAATTTCGTTACATAGTGTGTCAAAACGAGCTTAACCAT	531
Db	653	CAGATGGGCGCCCTGCACCATCTCCGGGTACCATGCGATCAACAACACATGACCCCT	712
Qy	532	GAGACTATCCACCTGAGAAAATCA-----CGACAGGGAACGCGACTGGC	576
Db	713	TATATCCTTCCACCGGGGACCATTTGTCTACAAGCTGATCCTTCATGTATCCAAAGGT	772
Qy	577	AAGAGATCCAGACTATCAACATGACTTTTACACTATGCTGTCTCGCCCATCTATC	636
Db	773	TACTTCGTCCAGACATATACCTTGCATCTTTCAACTATACGGGGCTGCACCGGTCGTG	832
Qy	637	TGGCTTATTTCTGTATCCCGCAGCAAACTATCCAGAGATTATCTGTGTTACAGATGTTGAT	696
Db	833	GTCTCTGACACACCCCTACACCTATATGATATGATATCACTGTGACCATGACGTGGAC	892
Qy	697	GGTGCATATGTCGTGATTAATCTACGAGGTGCAGATGGCGAACACAGACGCGGGCAGATC	756
Db	893	CGGACCTGTGGCGTGTGGAACTATCTGATTTCTGTCCAGGGCAGTGAACATTTCCAGCTA	952
Qy	757	CAGATCTCAGTGCATCGACGAGATGAGATGAGCTATTGTTCCAAAGGCTCGGAGCTCAGGGT	816
Db	953	GAAATGCGTCTTCTTGATGAGATGGATGCAGAAATTTGTGCCCTGTGAACAGGAATGAGGT	1012
Qy	817	ACTGTCACAATTCCTTCAGTCAAGCTATGGCAACTGGCGCGGCATATCTTCAACATC	876
Db	1013	CAACTTAAAGTGCACGGGTGCCCACTCTGGTGGCCTTACTGATGATGATGATTCAGCC	1072
Qy	877	CAGGTCAAAATTCGTGGGCTTCTAC-----GGCAGTGTATGTGCACACTCAAT	924
Db	1073	TACCTGTATCTTGTGAGGTGAGATGACATGACAAACCTGATGTGTGTGACTTCTTATACC	1132

```

QY 925 TTGGCTACGGGCGTGTCTACTCTCAAGTTCGCCGGTCACAATTCTTAATAATGAAG 984
Db 1133 CTCCTCTGTGGGATTTGAAACAGTGGCTGTACAAAGACAGTTCTCTAATAATGGAG 1192
QY 985 CCTTCTACTTACCGGTTTGGCAACATGAACACAGCATGTCGCAAGACAT 1044
Db 1193 CCTTCTACTTCCAAAGCGGTCAACCAAGCATGAGATTCAGATATCCGAGGAGAGCTTC 1252
QY 1045 GACCCAGCATACATGCTTCACATTCACATCAATAAATGATGGCAATTTCTTT 1104
Db 1253 GACTGGCTCTGTGTATTAAGATTTCAACCTCTCTGCTGGGGCAATTCCTTT 1312
QY 1105 CGGACTTCACTATCTTACCGCGAAGGTTCATGATTTGTCAGATCGAATGAAAT 1164
Db 1313 CGTACCAAGCACTATCTCTACTCTCGAGAGGTATCTTCACTCTGTGACCGATATGAAT 1372
QY 1165 GTCTGATGATGAACACCTGCGGTGTGTGACATTTGCTTATGAGGCGTATCTGAG 1224
Db 1373 GTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1421
QY 1225 AGTGTGCCCCCAACAATTTTACGCCAGATGCGATTAACATTAACCAAGAGGCCAC 1284
Db 1422 -----TTGGCAAGTGTCTCTTCGACAC 1447
QY 1285 AAGCAGGCGATTCGAGCTCATTTGCCGAGCAAAACATGCCAGTGTGTGATG 1344
Db 1448 CTAGAGGATGATGAGAGCTGTGTGCGAGGACAAATATCCTGCGGTGTATGTG 1507
QY 1345 TCTATTTGCAACGAGCCCGCATCTCATGAATGAGTGTGCGAATCTTGAGCCACTG 1404
Db 1508 TCTGTGCGCAATGAGCTGTCTCTCTGTAACCTGCGGATTTACTTCAAG----- 1561
QY 1405 ACCAATTGATCTGCTCACTTGATTCAACTGCGCTTATTCATTTGCTTAAGTGGCAG 1464
Db 1562 -----CTATATGCTTCCAAAGGCTTGAAGACATTCAGATTTCCCAAGAACTGGA-- 1609
QY 1465 GCGACATATCACTGATCGATCTGATCTGTTTGAATGATGATGATGATGATGATGAT 1524
Db 1610 GTTACAGCTCTTGTAAATACCAATGCCCGTACGTCGATGATGATGATGATGATGATGAT 1669
QY 1525 TTTCGATGATTTTCTCAACAGAGACCTTGAAGAGCAGAGCTCTTGAAGAGAG 1584
Db 1670 TTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1729
QY 1585 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1644
Db 1730 TTGAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1789
QY 1645 ACCCTTGAAGCTTCACTCTATCTGAGCTGCTTGAAGAGAGTTCAGATGACAA 1704
Db 1790 GCGCTCTGCGGGCTTCAATGAGATCCACCTGCAATGTCATGAGAGTACCAAGACCT 1849
QY 1705 ATGCTAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1758
Db 1850 CTCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1909
QY 1759 CATGTTGAACTTGTGCGATTTCAAGACCAATTTGGTATCATCTCGAGTGAAGCGTAC 1818
Db 1910 CTATCTGGAATTTTGTGATCTTCAATGAGAACCAATCACTGAGATTAACAGAAAC 1969
QY 1819 AAGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1878
Db 1970 AAGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2029
QY 1879 AGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1898
Db 2030 AGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2049

```

```

RESULT 6
CR593823 1377 bp mRNA 11near HTC 21-JUL-2004
LOCUS CR593823
DEFINITION full-length cDNA clone CS0DL001YM21 of B cells (Ramos cell line)

```

```

ACCESSION      CR593823
VERSION        CR593823.1 GI:50474630
KEYWORDS       HTC; CNSLT cDNA
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1377)
AUTHORS        Li, W.B., Gruber, C., Jesse, J. and Polayer, D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished
REMARK         Contact : Feng Liang Email : fliang@life.technet.com URL :
               http://fulllength.invitrogen.com/Invitrogen Corporation 1600
               Faraday Avenue
               (bases 1 to 1377)
REFERENCE      Genoscope.
AUTHORS        Direct Submission
TITLE          Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL        BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
               - Web : www.genoscope.cns.fr)
COMMENT         1st strand cDNA was primed with a NotI-Gli30(dt) primer. Five prime
               end enriched, double-strand cDNA was digested with Not I and cloned
               into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
               was normalized. Library was constructed by life technologies, a
               division of invitrogen.
FEATURES       Location/Qualifiers
               source
               1..1377
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="CS0DL001YM21"
               /tissue_type="B cells (Ramos cell line) Cot 25-normalized"
               /plasmid="pCMVSPORT_6"
ORIGIN
Query Match      9.6%; Score 183.2; DB 3; Length 1377;
Best Local Similarity 51.2%; Pred. No. 9.8e-46;
Matches 551; Conservative 0; Mismatches 468; Indels 57; Gaps 3;
QY 829 CCTCAGTCAGCTATGCGCAACCTGGCGCGCATATCTTACCAACTCCAGTCAATC 888
Db 145 CCGTACCTGATGACGAAACGCGCTGCTATCTGATTCATTTGAGAGGTGACATGCA 204
QY 889 GTGGGTTTACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 948
Db 205 CAGACGTACATGGGGCTGTGTCTGATCTTACACATCTCTGTGGGATCCGACGTG 264
QY 949 AAGGTGCGGGGTCAATTTCTTAATAATGAAGCTTTCTTACTTACCGGTTTGGC 1008
Db 265 GCTGTACACCAAGACCGATCTCTCATCATGAGGAACCTTTCTAATTTCCACGGTGTCAAC 324
QY 1009 AAACATGAAGCAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1068
Db 325 AAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 384
QY 1069 TTCAACTCATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1128
Db 385 TTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 444
QY 1129 GAAGAGGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1188
Db 445 GAGGAGTGTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 504
QY 1189 GTTGTCTGAACATTTGCTGATGAGCGTATCTGAGAGTGTGCCCAAAACATTTAG 1248
Db 505 GTGGGCTGGCGT-----G 519
QY 1249 CCAATGCGATTAACATTAACCAAGAGGCCCAAGCAGAGCATTCGTGAGCTCAT 1308
Db 520 CCGCAGATTTCTTCAACACCTTTCTGATCAACACATCAGTATGAGAGAAATGGTG 579
QY 1309 GCCGAGACAAACATCAGTGTGTCATGTGTCTATTGTCACAGAGCCGCACT 1368

```

```

DB 580 CGTAGGAGCAAGAACACCCCGCGTGTGATGTGTGTGCGCAACGAGCTGCGTCC 639
QY 1369 CATGAAGATGAGAGCTGCGGATCTTCGAGCCACTGACCAATTGATGCTGCACTGAT 1428
DB 640 CACCTTAGAATCTGTGGTACTACTTGAAGATGTATGCTACACCAAACTCTTGAGC 699
QY 1429 CCAACTGCGCCTATTACATTTGTCTAACGTTCGCGACGGCGCATATGAGGTGATCGGATC 1488
DB 700 CCTCTCCGGCTGTGACCTTT-----GTGAGCACTCTTACTTATGACAGACGAAAGGG 753
QY 1489 TCTGATCTGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1548
DB 754 GCTCCGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATG 813
QY 1549 GACCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1608
DB 814 CACCTGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
QY 1609 CACAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1668
DB 874 CAGAGCCCATTTATTTATGAGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 933
QY 1669 CTCGAGCTGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1728
DB 934 CCACCTGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 993
QY 1729 TTTGAT-----CGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1782
DB 994 CTGGATCAAAAACGCGAGAAATACGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1053
QY 1783 CAGACCACTTGGGTATCATCCGAGTACGAGTACAGAGAGAGAGAGAGAGAGAGAGAG 1842
DB 1054 ATGACTGACATGATCACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1113
QY 1843 CGAAAGCGAAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1898
DB 1114 AAGCAACCAAAAAGTGCAGAGGTTCTTTTGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1169

RESULT 7
LOCUS CD014093 1853 bp mRNA linear EST 21-OCT-2003
DEFINITION 90135266 Single gene library Homo sapiens cDNA, mRNA sequence.
ACCESSION CD014093
VERSION CD014093.1 GI:37777622
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1853)
Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,
Au-Young, J., and Stuve, L.L.
PCR isolation and cloning of novel splice variant mRNAs from known
drug target genes
Genomics 83 (4), 566-571 (2004)
CONTACT: Jin, P.
Incyte Corporation
3160 Porter Drive, Palo Alto, CA 94304, USA
Tel: 650 621 8639
Fax: 650 621 8965
Email: pj@incyte.com.
Location/Qualifiers
1..1853
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_11b="Single gene library"
/notes="Vector: pDrive Cloning Vector; RT-PCR was performed
using gene-specific primers flanking the open-reading
frame. PCR products were subcloned into pDrive Cloning
Vector and sequenced completely using M13 forward and

```

```

ORIGIN
Query Match 9.6%; Score 183.2; DB 6; Length 1853;
Best Local Similarity 51.2%; Pred. No. 1.1e-45;
Matches 551; Conservative 0; Mismatches 468; Indels 57; Gaps 3;

reverse primers. Sequencing gaps were closed by
re-sequencing using primers flanking the gapped areas."

QY 829 CCTCAGTCACTATGAGCACTGCGCGCGCATATCTTACCAACTCCAGGTCACATC 888
DB 554 CCGTACTTATGACAGCAAGCCCTGCTATCTGTATTCATGAGAGTGAAGCTGATCGA 613
QY 889 GTGGGTTCTAGCGCGCATGTAGTCGACACCTTCAATTGGTACAGGCGCTGCTATCTG 948
DB 614 CAGAGCTCACTGGGGCTGTGTGATCTTCACTCACTCCCTGTGGAGATCCGCACTGG 673
QY 949 AAGGTTCCGGGTCACAATCTTAAATGAAAGACCTTTCTACTTAAACCGGTTTGGC 1008
DB 674 GCTGTCAACAAGAGCCAGTCTCTCATTAATGAGAAACCTTTCTAATTTCCACGGTCAAC 733
QY 1009 AAACATGAAGACACAGCACTATGCGCAAGACATGACCCAGCATACATGTTTCAGAT 1068
DB 734 AAGCATGAGAGATGCGGACATCCGAGAGAGGCTTGCATGCGCCGCTGCTGTGAAGAGC 793
QY 1069 TTCCAACTCATGAAATGATGAGAGCAATTTCTTTTGGACTTCACTATCTTACGGG 1128
DB 794 TTCAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 853
QY 1129 GAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1188
DB 854 GAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 913
QY 1189 GTTGTGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1248
DB 914 GTGGGCTGCGCT-----G 928
QY 1249 CCAGATGCGATTAACGATTAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308
DB 929 CCGCACTTCTTCAACAACCTTTCTCTGATACACATGATGATGATGATGATGATGATGATG 988
QY 1309 GCCCGAGCAAAAACCATGCGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 1368
DB 989 CGTAGGAGCAAGAACACCCCGCGTGTGATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1048
QY 1369 CATGAAGATGAGAGCTGCGGATCTTCGAGCCACTGACCAATTGATGCTGCACTGAT 1428
DB 1049 CACCTAGATCTGTGGTACTACTTGAAGATGTGATGCTCACCAAACTCTTGAGC 1108
QY 1429 CCAACTGCGCCTATTACATTTGTCTAACGTTCGCGACGGCGACATATCAGCTGATCGGATC 1488
DB 1109 CCTCTCCGGCTGTGACCTTT-----GTGAGCACTCTTACTTATGACAGACGAAAGGG 1162
QY 1489 TCTGATCTGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1548
DB 1163 GCTCCGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATG 1222
QY 1549 GACCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1608
DB 1223 CACCTGAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1282
QY 1609 CACAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1668
DB 1283 CAGAGCCCATTTATTTATGAGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1342
QY 1669 CTCGAGCTGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1728
DB 1343 CCACCTGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1402
QY 1729 TTTGAT-----CGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1782
DB 1403 CTGGATCAAAAACGCGAGAAATACGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1462
QY 1783 CAGACCAACTTGGTATCATCTGAGTACGAGTACAGAGAGAGAGAGAGAGAGAGAGAGAG 1842

```

Db 1463 ATACTGACAGACGACGACGAGAGTGCTGGGGAATTAAGGAGATCTTCACTGGCAG 1522

Qy 1843 CGAAGCGCAAGCGCGGACGCTACTAGTTTGAGGGCAAGGTGACGATGATTTGATTA 1898

Db 1523 AGCAACCAAAAGTACGCGCTTCTTTGCGAGAGATCTGGAAGATTGCCAA 1578

RESULT 8
BX745933 857 bp mRNA linear EST 18-NOV-2003
BX745933/c XGC-gastrula Xenopus tropicalis cDNA clone Tgas066d19 3',
LOCUS mRNA sequence.
DEFINITION BX745933.1 GI:38418673
VERSION BX745933
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 857)
REFERENCE 1 Croining,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
AUTHORS Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL Unpublished (2003)
COMMENT Contact: Croining MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: Tgas066d19.q1k17
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron W. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pcsi107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pcsi107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue.
location/Qualifiers
1..857
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="Tgas066d19"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_id="XGC-gastrula"
/note="Vector: pcsi107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pcsi107 with EcoRI at the 5' end and NotI at the 3'
end."

ORIGIN

Query Match 9.5%; Score 180.4; DB 5; Length 857;
Best Local Similarity 54.4%; Pred. No. 6.4e-45;
Matches 444; Conservative 0; Mismatches 321; Indels 51; Gaps 2;

Qy 907 GTATGCAACCCATTTGGCTAGCGCGCGCTACTGTCAAGTTGCCGGTCA 966

Db 789 GTGGAAGACATTAATTTGGCTAGCGCGCGCTACTGTCAAGTTGCCGGTCA 966

Qy 967 TTCTTAATAATGAAGCCTTCTACTTACCGGTTTGGCAACATGAAGACACCA 1026

Db 729 TTCTTATCAATGAAGAACATTTCTACTTCCATGTGTCAACAAATGAAGACAT 670

Qy 1027 GTACGCGGAAGGACATGACCCAGATATGTTCAAGTTGCCAATCATGAATGG 1086

Db 669 GTCAAGGAAAGGACATGACCTGCTCAATTTGAAGATTGAATCTGTGAAGTGG 610

Qy 1087 ATTGAGCAATTTCTTTGCGACTTCAACATCTTACCGGGAAGGTATGATTTCC 1146

Db 609 CTGTGCTAATCTCTTCGCGACACGACATTAATCTTATGACAGAAATCATGACCTT 550

Qy 1147 GCAGATGGAATGGAATTTGCGATGCGATGAAGAAACCTCCGTTGTGTGAACATGGCC 1206

Db 549 TGTGATTAATATGACATTTGTGTGATTAAGAAATGCTCGAGATGTATTA 497

Qy 1207 TTGATGGCGTATCTGAGAGTGTGCCCCAACAATTTACCCAGATGCGATTAAGAT 1266

Db 496 -----ATACCCGAAAGTTTGGAAAC 475

Qy 1267 AAAACCGAAGGCGCCACAGACGAGATTCGTGAGCTCATTTGCCGAGACAAACCAT 1326

Db 474 CAATCTTTAAACACCATTTTAATAGTTATGAGAGAGTTGTGCGAGGATTAACCGG 415

Qy 1327 GCCAGTTGTGATGATGTGTCTATTTGCAACGAGCCGCGATCATGAAGATGAGCTGC 1386

Db 414 CTTCCGTTGTGATGATGTGTCTGTTGCAATGACACGATCCAGCTCCCTGTGCTGG 355

Qy 1387 GAATACTTGAGCCACATGACCAATTTGACTGCTCACTTGAATCCAACTGCGCTATTACA 1446

Db 354 TATTAATTTAAACCGTATTTGTTACCTAAGCAACTTGACCCACCCAGACAGTGACA 295

Qy 1447 TTTGCTAACGTGCGACGCGACATATCATGCTGATGCGATCTGATCTTTGATGTC 1506

Db 294 TATGTTTCCATGCTAATCATGAAACATGACCAAGTGCACCTTATG-----TGAATGA 241

Qy 1507 AGTTGCAATTAATCGGTATTTGGATGTATTCGAAACAGAGACCTTGAGAAAGCAGAG 1566

Db 240 ATTGTGTAAACAGTTACTTTCTTTGATCATATGCGCGGACCTGGAAGTTATCCA 181

Qy 1567 GCAGCTCTTGAAGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1626

Db 180 CTCCAACTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 121

Qy 1627 ACCGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1686

Db 120 AGTGAATATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 61

Qy 1687 GAAGAGTTCCAAATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1722

Db 60 GAAAGATCCAGAGAGTGTGTGAAGAACTACCAT 25

RESULT 9
Cl679241 811 bp DNA linear GSS 09-JUL-2004
LOCUS PRI0125C_C10.2 - PRI0125C.BR (811) Mixed stage fosmid library of P.
DEFINITION pacificus var. California Pristionchus pacificus genomic
survey sequence.
ACCESSION Cl679241
VERSION Cl679241.1 GI:50185841
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Nemodiplogasteridae; Pristionchus.
1 (bases 1 to 811)
REFERENCE 1 Strinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
AUTHORS Appab: an AcedB database for the nematode satellite organism
JOURNAL Pristionchus pacificus
COMMENT Nucleic Acids Res. 32 (1), D421-D422 (2004)
Contact: Sommer R.J
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@uebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.
location/Qualifiers

source
 1. .811
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="California"
 /db_xref="taxon:54126"
 /clone_lib="Mixed stage fosmid library of P. pacificus
 var. California"
 /note="Vector: pBfi105-5 Fosmid vector"

Query Match 9.4%; Score 178.8; DB 9; Length 811;
 Best Local Similarity 53.8%; Pred. No. 2e-44;
 Matches 444; Conservative 0; Mismatches 367; Indels 15; Gaps 3;

QY 289 ATTACAGACCATGGAGATGGGTTTCTATCAGCGAGGTCATTTGCCCAAGGCTGG 348
 DB 1 ATTCTGATTTATTTGGGCAACGTTCTGTATCAGCGAGGTCATTTATCCGAAAGGTTGG 60
 QY 349 TCTCAGAGGATATCTGTCGAGCGCATTCGCTACGACCATGGTCGATCTATGTC 408
 DB 61 GAGGCGCAGGATTCGTCGTCGCTTCGATGCGGTCACTATACGCAAAAGTGGGTC 120
 QY 409 AACACCGGCTTGTGCGAGCATGTGGCGGCTATACCTTTTGAAGCGACGTCAT 468
 DB 121 AATTAATCAGAAATGATGAGCATCAGGCGGCTATACGCTTTGAAGCGGATGTCACG 180
 QY 469 GAATTAATGCGCGCGGAGAAATTCGCTTACGATGGTGTCAACAGAGCTTACC 528
 DB 181 CCTATGTTATTTCCGGAAGATGATCATCGTTGTGTGAACAAGATGAGAAC 240
 QY 529 CATGACATATCCACCTGGAAAAATCAGACAGGAAACGCACTGGCAAGAAATCCAG 588
 DB 241 TGGCAGACTATCCGCGGGAATGTGATTAACG---ACGAAAACGCAAGAAAAGCAG 297
 QY 589 ACCTATCAACATGACTTTTACAATGCTGTGCTGCGCCGATATCTGGCTTTATCT 648
 DB 298 TCTTACTTCATGATTTCTTAACTACGCGGAGATCCATGCGAGCATGATCTCTACCC 357
 QY 649 GTATCCAGCAATATCCAGATATTAATGATGATGATGATGATGATGATGATGATGAT 708
 DB 358 AGCGCAACACTGGGTTGACGATATCAGTGTGATGATGATGATGATGATGATGATGAT 417
 QY 709 CTGATTAATCTACAGGATCGAAGTGGCAACCAAGCAGCGGAGATCCAGATCTCAGTC 768
 DB 418 C-----AGCGCTGTGTGATCTGGCAGGTGTGTGCAATGTGATGTGATGATGATGAT 471
 QY 769 ATCGACGAGATGAGACTATTTGTGCAAAAGGCTCTGGAGCTCAGGATCTCAGCAAT 828
 DB 472 CGTATGCGGATCAACAGGTGTGCAACCTGCAAAAGCAGCGGAGCTTTGCAAGTC 531
 QY 829 CCTCATGTAAGTATGAGCACTGGCGCGCATCTCTACCAACTCCAGGTCACATC 888
 DB 532 GTGATATCCGACTCTGGCAATGAGGTGAAGTTATCTTATGAACTGTGTCACAGCC 591
 QY 889 GTGGGTTCTAGCGGATGATGCAACCTTACATTTGGCTACGGGCGGTCTACTGTC 948
 DB 592 AAAAGCAGACAGAGT-----GATATCTACCGCTGGCGGTCCGATCCGCTCAGTC 645
 QY 949 AAGGTTCGCGGCTCACAATTTCTTAATATGAAAGCTTTTCTTACCTGCTTTTGGC 1008
 DB 646 GCGATGAAAGGCGCAAGTCTGATCAACCAAAACGCTTTTCTTACTGCTTTTGGC 705
 QY 1009 AAACATGAGACACAGAGTATGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGT 1068
 DB 706 CGTCAATGAAAGTGGGATTTGGCGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGT 765
 QY 1069 TTCCAATCATGAAATGATGAGCAAAATCTTTTGGACTTCAC 1114
 DB 766 CAGGATTAATGAGCTGAGATTTGGGCAACTCTACCGTACCTGCG 811

RESULT 10
 CL463714/c

LOCUS CL463714 908 bp DNA linear GSS 31-MAR-2004
 DEFINITION SAIL_119_H09.v1 SAIL Collection Arabidopsis thaliana genomic clone
 SAIL_119_H09.v1, genomic survey sequence.
 ACCESSION CL463714
 VERSION CL463714.1 GI:45866619
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 (bases 1 to 908)
 Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
 Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
 Bullis,D., Snell,J., Miguel,T., Hutchinson,D., Kimmery,B.,
 Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
 A high-throughput Arabidopsis reverse genetics system
 Plant Cell 14 (12), 2985-2994 (2002)
 JOURNAL MEDLINE
 PUBMED 22356987
 COMMENT Contact: Sessions A
 Applied Trait Genetics
 Syngenta Biotechnology Inc.
 3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
 Email: allen.sessions@syngenta.com
 ABRC Stock Number CS805872; T-DNA left border flanking sequences of
 Syngenta Arabidopsis Insertion Library (SAIL) lines are available
 through the Arabidopsis Biological Resource Center (ABRC).
 Sequences represent a pool of amplified genomic regions and not
 single contiguous sequences.
 Class: TDNA tagged.
 Location/Qualifiers

FEATURES
 source 1..908
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Columbia"
 /db_xref="taxon:3702"
 /clone_lib="SAIL_119_H09.v1"
 /clone_lib="SAIL_Collection"
 /note="T-DNA left border sequences were isolated using a
 modified TAIL-PCR strategy"

ORIGIN

Query Match 9.2%; Score 175.4; DB 9; Length 908;
 Best Local Similarity 64.3%; Pred. No. 2.5e-43;
 Matches 263; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1480 GATCGATCTGATCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1539
 DB 715 GATACATCAGGATCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 656
 QY 1540 CAACACAGAGACCTTGAGAGAGACAGGACGCTTTGAAAAGAGAGCTGATGATGAGCA 1599
 DB 655 CAAGGCGGATTTGGAAACGAGAGAGATGATGAAAAGAACTTCTGGCTTGGCAG 596
 QY 1600 GAGAAATTCACAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1659
 DB 595 GAGAACTGATCAGCGGATTTATCATCAGGATGATGATGATGATGATGATGATGATGAT 536
 QY 1660 CACTTATCTCGGATGCTGCTTGGAGCGAAGTTCCAAAGTACAAATGCTAGCATGTAC 1719
 DB 535 CACTGATGATCTCCAGCATGAGTGAAGTATCACTGTCATGATGATGATGATGATGAT 476
 QY 1720 CATGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1779
 DB 475 CACCGGCTTTGATGCTCAGCGCGCTGTCGCGGAAACAGATGAGAAATTTCCGCGCAT 416
 QY 1780 TTCCAGAACCACTGGGATTCATCCAGATGAGCGGTAACAAGAGGAGTTTTCACCGGT 1839
 DB 415 TTTCGACCTCGCAGAGCATATTTGCGGCTTGGCGGTAACAAGAGGATCTTCACTCCG 356
 QY 1840 GACCGAAACCAAGCGGACGCTCATTAATTTGAGGCAAGGTGACTA 1888

JOURNAL Plant Cell 14 (12), 2985-2994 (2002)
 MEDLINE 22356987
 PUBMED 12468722
 COMMENT

Contact: Sessions A
 Applied Trait Genetics
 Syngenta Biotechnology Inc.
 3054 Cornwell Rd., Research Triangle Park, NC 27709, USA
 Email: allen.sessions@syngenta.com
 ABRIC Stock Number CS801415; T-DNA left border flanking sequences of
 Syngenta Arabidopsis Insertion Library (SAIL) lines are available
 through the Arabidopsis Biological Resource Center (ABRC).
 Sequences represent a pool of amplified genomic regions and not
 single contiguous sequences.
 Class: TDNA tagged.
 Location/Qualifiers

FEATURES
 source
 1..907
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Columbia"
 /db_xref="taxon:3702"
 /clone="SAIL_28_D03.v1"
 /clone_1ib="SAIL Collection"
 /note="T-DNA left border sequences were isolated using a
 modified TAIL-PCR strategy"

ORIGIN

Query Match 8.7%; Score 164.8; DB 9; Length 907;
 Best Local Similarity 59.2%; Pred. No. 5.7e-40;
 Matches 302; Conservative 0; Mismatches 202; Indels 6; Gaps 1;
 Oy 1125 CCGGAGAGATCATGATTTGCGAGATGGAATTTGTCATGATGAACACC 1184
 Db CCGGGGGGGGGCGCTTCCCATATATGACATGCGATGCTGATGATGAACCTCG 769
 Oy 1185 TCCCGTGTCTGAC-----ATTGCTTGTATGAGCGATCTGAGATGTCGCCACA 1238
 Db 768 TCCCTGCGGCTTAACTCTCTTTAGGCAATGCTTGAAGCGGGCAACAGCCGAAGA 709
 Oy 1239 AACATTAGCCAGATGATTAAGATTAACCCAGAGGCCACAGAGGCGATTCG 1298
 Db 708 ACTGTACAGCGAAGAGGAGTCAACGGGAACTCAGCAAGCGCACTTACAGCGATTA 649
 Oy 1299 TGAGCTCATTTGCCGAGCAAAAACCATGCGAGTGTTCATATGCTTATTTCCAAACA 1358
 Db 648 AAGAGCTATAGCCGCGTGAACAAACCAACGCGTGTGATGTGAAGTATTCACCAACA 589
 Oy 1359 GCCCGCATCTCATGAAGATGAGCTGCGGAATTCCTGAGCCACTGACCAATTTGA 1418
 Db 588 ACCGGAATCCCGTCCGCAAGTGCACGGGAATATTTCCGCCACTGCGGGAACCAACGCG 529
 Oy 1419 TCACTTATTCACATCTGCGCTTATTAATTTGCTAACGTCGCGACCGACATATCAGCT 1478
 Db 528 TAAACTCGACCCGACCGCTCCGATCACTGCGCAATGTAATGTTTGGCAGCGCTCAC 469
 Oy 1479 GGATCGATCTCTGATCTGTTTATGATGATGATGATGATGATGATGATGATGATGAT 1538
 Db 468 CCAATACATCAGGATCTCTTTGATGTGCTGCTGACCGTTATTTACGAGATGTGTG 409
 Oy 1539 TCAACAGAGACCTTGAAGAGCAGAGGAGCTCTTGAAGAGAGCTGATGATGAGCA 1598
 Db 408 CCAAGCGGCGATTTGTAGCGGCAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 349
 Oy 1599 AGAGAAATTCACAGCGCGATGCTCATGAC 1628
 Db 348 GAGAAATCTGCATCAGCGCATTTATCATGTC 319

RESULT 13
 CD503098 1138 bp mRNA linear EST 12-JUN-2003
 LOCUS CD503098
 DEFINITION CDA60-D07.xid.c SHGC-CDA Gasterosteus aculeatus cDNA clone
 CDA60-D07 5', mRNA sequence.
 ACCESSION CD503098

VERSION CD503098.1 GI:31433163
 KEYWORDS EST.
 SOURCE Gasterosteus aculeatus (three spined stickleback)
 ORGANISM Gasterosteus aculeatus

REFERENCE
 AUTHORS 1 (bases 1 to 1138)
 Kingley, D.M., Petchel, C., Balaband, S., Grimwood, J., Dickson, M.,
 Schmutz, J., and Myers, R.M.
 TITLE Expressed sequence tags from Gasterosteus aculeatus
 JOURNAL Unpublished (2003)
 COMMENT Contact: Kingley, DM
 HMI and Department of Developmental Biology
 Stanford University School of Medicine
 Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
 Tel: 650 725 5954
 Fax: 650 725 7739
 Email: kingley@cmgm.stanford.edu
 Place: 60

FEATURES
 source
 1..1138
 Location/Qualifiers

/organism="Gasterosteus aculeatus"
 /mol_type="mRNA"
 /strain="Salinas river, CA"
 /db_xref="taxon:69293"
 /clone="CDA60-D07"
 /sex="mixed male and female"
 /tissue_type="heads and internal organs combined"
 /dev_stage="adult"
 /clone_1ib="SHGC-CDA"
 /note="Vector: lambda ZAP Express/PBK-CMV. Site 1: EcoRI
 (5' adaptor); Site 2: XhoI (3' linker primer). The mixed
 organ cDNA library was generated using the ZAP-cDNA method
 by Stratagene. First strand cDNA synthesis was primed with
 a 50 bp linker primer containing an oligo dt sequence
 preceded by a synthetic XhoI site. 5 prime adaptors were
 used containing an EcoRI cohesive end. The finished cDNAs
 were inserted in to the ZAP express vector
 unidirectionally in the sense orientation with respect to
 the lacZ promoter of PBK-CMV. An amplified library was
 prepared from approximately 3 million primary clones in
 the lambda ZAP Express vector. In vivo excision was then
 used to generate individual PBK-CMV phagemid clones for
 EST sequencing."

ORIGIN

Query Match 8.3%; Score 158.8; DB 6; Length 1138;
 Best Local Similarity 52.8%; Pred. No. 5e-38;
 Matches 502; Conservative 0; Mismatches 387; Indels 61; Gaps 5;
 Oy 912 CGACACTTCAATTTGGCTACGGCGCTGCTACTGTCAAGTTGCCGGTCAATTTCTT 971
 Db 73 CGACGTGATCACTCTACAGAGTGGCATCGCACGGTGAAGTTACCAAGACCAAGTTCC 132
 Oy 972 AATAATGAAAGCTTTTCTTACCGGTTTGGCAACATGAAGACAGCAGTACG 1031
 Db 133 CATCAACAAAGAGCCCTTCTACTTCCAGAGTAATTAACAGAGACTCTGATATTCG 192
 Oy 1032 TGGCAAGAGCATGACCAAGCATACATGATTCAGATTTCAACTGATGAATGATTCG 1091
 Db 193 AGCAAAAGGCTGGATGGCCCTCATGGTGAAGACCTTTAACTTTGAAGTGGTTGG 252
 Oy 1092 AGCAATTTCTTTTGGACTTCACTACTCTTAACGGGAAGAGTGAATTTGGCAGA 1151
 Db 253 GGCCAATCTGTTCCGACAGCACTACCTTATGAGAGAGATCTCGAATGTGTGA 312
 Oy 1152 TGAATGGAATTTGTGATGATGAACACTGCGCTTGTGTGAACATTCCTTGA 1211
 Db 313 CCGCCATGCGATCTGTGATGAAGAGATGCGCGGCGGTGG----- 354

QY 1212 GGGCGTATCTGAGAGTGTCGCCCAACAACTTTACGCGAGATGCGATTACGATTAAC 1271
 DB 355 -----CATTAAGACATTTGCGAGTTTGGAAACGCC 385
 QY 1272 CCNAGAGGCC--CAACAGCAGCGATTCGTAGCTCATTTGCCGAGACAAAACATATGCC 1329
 DB 386 TCCCTTAACCACTACCTGCTGTCTGTATGAGACGACTGCTGATGCTGGGAGCAAGAACATATCC 445
 QY 1330 AGGTGTTCATGTGTGTATTTGCAACGAGCCCGCATCTCATTAAGATGAGCTGCCGA 1389
 DB 446 TCTGTGTGTATGT 505
 QY 1390 TACTTGAGGCGCATGACCAATTGACTCGTCAACTTATTCACACTGCGCCCTTATTCATTT 1449
 DB 506 TATTTCAAAACCTTGATTAATAACATACCAAGATTTGATTCACACCGGCGCTCATTTT 565
 QY 1450 GCTAACGTCGCGCAGCGGACATATCATCTGATTCGATCTCGATCTGTTTATGTCAGT 1509
 DB 566 ATCACAAGACAG-----TAACTATGCGAGGATTAAGGGGCTCCCTACGTCGACGTCATC 619
 QY 1510 TGCATTAATCGGTATTTGCGATGTGTCTCAACAGAGACCTTGAGGAAACAGAGGCA 1569
 DB 620 TGGTAAACAGATTACTTCTCTGTACATGACCCGGGCGACCGGAGTGTATCCCATC 679
 QY 1570 GCTCTGAAAAGAGAGCTGATGATGAGGAGAAATTCACAGCCGATGTCATGAC 1629
 DB 680 CAGCTCAACATCAGTTGATGAGATGATGAGAAAGTACAGAAACCCATCATCCAGAGC 739
 QY 1630 GAATATGTGTCAGATACCTTTGACAGCCCTTCACTCTATCTCGAGACTGCTTGAGGGA 1689
 DB 740 GAATACGAGAGCGATGCGGTGCGGGGCTTCAAGTATCAACCGGTATTTTACTGAG 799
 QY 1690 GAGTTCAGAGTCAATGCTAGACATGTACATGATGAGTGTATTC-----GCATTGAGT 1744
 DB 800 GAGTACCAAGATTTGCTCTGAGAGCTACCAACAGTGTTCGACGAGAAAGAGAGT 859
 QY 1745 CGATGCGAGGAGCATGTTTGAACCTTCCGCAATTTCCAGACCAATTTGGTATCATTC 1804
 DB 860 ACGTATCGGCGAATCATCTGGAACCTTTGACAGCTTCATGACCAACAGGATCATGTC 919
 QY 1805 GAGTACGAGTAAACAAAGAGGTGTTTTCACCCGCTGACCGAAAGCCCAAG 1854
 DB 920 GCGTGTGTGGAGAACAG-AGGCTGTCTTCAAGCAGAAAGGACGCAAG 968

 RESULT 14
 CFS21612 740 bp mRNA linear EST 10-SEP-2003
 LOCUS AGENCOURT 15508970 NICHG_XGC_Kid1 Xenopus laevis cDNA clone
 DEFINITION IMAGE:7012148 5', mRNA sequence.
 ACCESSION CFS21612
 VERSION CFS21612.1 GI:34572485
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 Xenopodinae; Xenopus; Xenopus.
 1 (bases 1 to 740)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Igor Dawid
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed By: The I.M.A.G.E. Consortium (LIMU)
 DNA Sequencing By: Agencourt Bioscience Corporation
 Clone distribution: NCI-CGAP clone distribution information can be

Found through the I.M.A.G.E. Consortium/LIMU at:
<http://image.llnl.gov>
 Plate: LLMJ4721 row: n column: 18
 High quality sequence stop: 670.
 Location/Qualifiers
 1. 740
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:7012148"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Kidney; Vector: pCMV-Sport6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.2 kb. Constructed by Life
 Technologies. Note: This is a Xenopus Gene Collection
 (XGC) library."

ORIGIN

Query Match 8.2%; Score 155.4; DB 7; Length 740;
 Best Local Similarity 53.1%; Pred. No. 5, 1e-37;
 Matches 416; Conservative 0; Mismatches 316; Indels 51; Gaps 2;

QY 907 GTATGACACCTCAATTTGGCTACGGCGCGCTACTGTCAAGTTGCCGGGTCAAA 966
 DB 9 GTGGAAGATTTTACGTTTGGCTGTGGAATTAAGACTGTGCACTCTCTGAGACCA 68
 QY 967 TTCTTAATTAATGAAGCTTTTCTACTTTACCGGTTTGGCAACATGAAGACAGCA 1026
 DB 69 TTCTCATATATGAAGAACATTTCTACTTCATGATGTGCAACATGAGATATAT 128
 QY 1027 GTACGCGGAAAGGACATGACCCAGCATACATGTTTCAACTATGAATATG 1086
 DB 129 GTCAAGGTAAAGGACATGACTGTCTACTGTCAATTAAGATTTTATCTGTTGAAGT 188
 QY 1087 ATTGAGCAAAATCTTTTGGACTTTCACATATCTTACGCGGAGGATGATGATTC 1146
 DB 189 CTGTGTCTAATCTCTTCCGAAACAGCATTTATTCATATGCAAGAAATCATGAGCTT 248
 QY 1147 GCAGTCGAATGAATTTGTGTATGATGAAGAACATCTGCTGTGAACATTTGCC 1206
 DB 249 TGTGATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301
 QY 1207 TTGATGGCGGTATCTGAGAGTGTGCCCCCAAAACATTTACGCAATGAGATTAACAT 1266
 DB 302 -----ATACCTTAAGTTTGGAAAC 323
 QY 1267 AAAACCAAGAGGCCCAACAGAGCGATTCGTAGCTCATTTGCCGAGACAAAACAT 1326
 DB 324 CAATCTTTAAACCAACATTTAATATGATGAGAGAGTGTGTGCGGAGTAAAGACCG 383
 QY 1327 GCCAGTGTGTCAATGTGTCTATTTGCAACGAGCCCGCATCTCATGAATGAGCTGCC 1386
 DB 384 CTTTCGTTGTCTATGT 443
 QY 1387 GAATCTTTCAGGCACTGACCAATTTGACTGTCACTTATCAACTGCGCCCTATTTCA 1446
 DB 444 TATTTATTTTAAACCGTGAATGATGATCACTAAGCAACTTATCCACAGACCGGTGCA 503
 QY 1447 TTGTAAAGTGTGCGGAGGAGCATATACGCTGATCGGATCTGTATCTGTTGATGTC 1506
 DB 504 T-----AGTTTCTATATCTATGATGATGATGATGATGATGATGATGATGATG 557
 QY 1507 AGTTGATTAATCGGTATTTCCGATGATGATTTCTCAACAGAGACCTTGAGAGAGAG 1566
 DB 558 ATCTGTGTAATAGTATCTTTCTGTATATGATGATGATGATGATGATGATGATGATG 617
 QY 1567 GAGGCTTTGAAAGAGAGTGTATGATGAGCAAGAAATTTCAACAGCCGATGATGATG 1626
 DB 618 CTCCAATCAATTAATCAATTTGCAAGGATGATGAGATGATCAAAAGCTATGATACAG 677
 QY 1627 ACCGAATATGTGAGATACCTTGACAGGCTTCACTATCTCTGAGATGCTTTGAGC 1686

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: March 22, 2005, 18:23:41 ; Search time 4458 Seconds

(without alignments)
5413.357 Million cell updates/sec

Title: US-10-757-093-4

Perfect score: 3354

Sequence: 1 MKFLTGLSLAAPSIGTP.....RKPKAAHSLRAWTSIDKN 634

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=alp
-Q/cgn2_1/USPTO.spool_p/US10757093/runat_18032005_164456_27770/app_query.faeta_1.775
-DB=EST -QFMT=faetap -SUFPLX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0
-UNIT=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=prc -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10757093.@CEN_1_1.5180.@runat_18032005_164456_27770 -NCPU=6 -ICPU=3
-NO_MMWP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -XGAPOP=6
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:.*
1: gb_est1:.*
2: gb_est2:.*
3: gb_hic:.*
4: gb_est3:.*
5: gb_est4:.*
6: gb_est5:.*
7: gb_est6:.*
8: gb_gsa1:.*
9: gb_gsa2:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1267	37.8	2473	3	AY321342 Rattus no
2	1253.5	37.4	2095	6	CD014094 90135027
3	1175	35.0	1995	6	CD014092 90134967
4	1168.5	34.8	2274	3	AK041058 Mus muscu
5	1045.5	31.2	1853	6	CD014093 90135266
6	946	28.2	1377	3	CR593823 full-len
7	803	23.9	1124	6	CD503076 CDA60-C07
8	797.5	23.8	1051	5	BX363460 BX363460
9	797	23.8	921	9	CL486845 SATL_443

10	770.5	23.0	1138	6	CD503098	CD503098 CDA60-D07
11	726.5	21.7	857	5	BX745933	BX745933 BX745933
12	712.5	21.2	811	9	CL679241	PR10125C
13	705.5	21.0	906	5	BO941196	AGENCOURT
14	689	20.5	1055	4	BM557676	AGENCOURT
15	683	20.4	877	7	CF406520	CH3#042_G
16	681.5	20.3	767	7	CF255373	mdvnl27_c
17	674.5	20.1	740	7	CF521612	AGENCOURT
18	671.5	20.0	816	7	CK467358	938649 MA
19	671.5	20.0	878	7	CN155220	942501 MA
20	661.5	19.7	914	6	CB203472	AGENCOURT
21	656	19.6	1063	5	BX401772	BX401772 BX401772
22	649.5	19.4	915	5	BO678153	AGENCOURT
23	648.5	19.3	846	6	CA453907	AGENCOURT
24	647.5	19.3	802	5	BU357212	603474094
25	643	19.2	733	7	CN254047	170006000
26	637	19.0	736	5	BP454056	BP454056 BP454056
27	633.5	18.9	791	5	BU239978	BU239978 603323761
28	630.5	18.8	689	7	CR442804	CR442804 CR442804
29	622.5	18.6	725	5	BU315788	603851444
30	621	18.5	566	7	CV169266	rmexr1_01
31	620.5	18.5	900	6	CA489317	AGENCOURT
32	609.5	18.2	799	5	BU366987	BU366987 603585124
33	609	18.2	771	4	BG121498	602352830
34	605.5	18.1	731	7	CV105482	AGENCOURT
35	601.5	17.9	780	7	CO559438	AGENCOURT
36	601	17.9	690	7	CK836522	4061270 B
37	591	17.6	898	5	BU179563	AGENCOURT
38	590	17.6	962	5	BQ684422	AGENCOURT
39	587.5	17.5	733	7	CF748482	UI-M-HJ0-
40	585.5	17.5	757	4	B1851705	603377940
41	583.5	17.4	847	5	BX621253	BX621253
42	582	17.4	925	5	BO679162	AGENCOURT
43	578.5	17.2	624	4	BI445636	deeb305
44	577.5	17.2	673	4	CB841568	CB841568 M5E-2121
45	573	17.1	857	6	CB559443	AGENCOURT

ALIGNMENTS

RESULT 1
AY321342
LOCUS
DEFINITION Rattus norvegicus Ac2-223 mRNA, complete cds.
ACCESSION AY321342
VERSION AY321342.1 GI:32527744
KEYWORDS
SOURCE
ORGANISM Rattus norvegicus (Norway rat)
HTC.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 2473)
XU,C.S., LI,W.O., LI,Y.C., WANG,L., WANG,S.F., HAN,H.P., WANG,G.P.,
CHAI,L.Q., YUAN,J.Y., YANG,K.J., YAN,H.M., CHANG,C.F., ZHAO,L.F.,
MA,H., SHI,J.B., RAHMAN,S., WANG,Q.N. and ZHANG,J.B.
Liver regeneration after PH
Unpublished
2 (bases 1 to 2473)
XU,C.S., LI,W.O., LI,Y.C., WANG,L., WANG,S.F., HAN,H.P., WANG,G.P.,
CHAI,L.Q., YUAN,J.Y., YANG,K.J., YAN,H.M., CHANG,C.F., ZHAO,L.F.,
MA,H., SHI,J.B., RAHMAN,S., WANG,Q.N. and ZHANG,J.B.
Direct Submission
Submitted (11-JUN-2003) Henan Bioengineering Key Lab, Henan Normal
University, NO. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R.
China

FEATURES

source
1..2473
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
41..2467
/note="liver regeneration related protein LRRC134"

QY 585 GUGUUNhVAlTTPAnPhEAlAaRPhEglnThrAsnLeuGlyLeileArgValaAr 604
 DB 1904 GGAGACTATCTGGAAATTTGCTAGCTTATGACAAACAGTCCACCACGAGAAACA 1963
 QY 605 GUAhNLYuLgVAlPheThrARgARgLYuProlYAlAlaAlaNIsserLeu 624
 DB 1964 GGAAACAAAGGGATCTTCACTCGACAGAAACCCCAAGATGGACGCTTCATTGG 2023
 QY 625 ArgAlaArgTTrpThrsertle 631
 DB 2024 CGAGAGAGATCTGAGAGATT 2044
 RESULT 2
 LOCUS CD014094 2095 bp mRNA linear EST 21-OCT-2003
 DEFINITION 90135027 Single gene library Homo sapiens cDNA, mRNA sequence.
 ACCESSION CD014094
 VERSION CD014094.1 GI:37777623
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2095)
 JIN, P., FU, G. K., WILSON, A. D., YANG, J., CHEN, D., HAWKINS, P. R.,
 AU-YOUNG, J. and STUVE, L. L.
 PCR isolation and cloning of novel splice variant mRNAs from known
 drug target genes
 Genomics 83 (4), 566-571. (2004)
 CONTACT: Jin, P.
 Incyte Corporation
 3160 Porter Drive, Palo Alto, CA 94304, USA
 Tel: 650 621 8639
 Fax: 650 621 8965
 Email: pj@incyte.com.
 location/Qualifiers
 FEATURES
 source 1..2095
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="Single gene library"
 /note="Vector: pDrive Cloning Vector; RT-PCR was performed
 using gene-specific primers flanking the open-reading
 frame. PCR products were subcloned into pDrive Cloning
 Vector and sequenced completely using M13 forward and
 reverse primers. Sequencing gaps were closed by
 re-sequencing using primers flanking the gapped areas."
 ORIGIN
 Alignment Scores:
 Pred. No.: 4..21e-129 Length: 2095
 Score: 1253.50 Matches: 266
 Percent Similarity: 57.51% Conservative: 90
 Best Local Similarity: 42.97% Mismatches: 187
 Query Match: 37.37% Indels: 77
 DB: 6 Gaps: 13
 US-10-757-093-4 (1-634) x CD014094 (1-2095)
 QY 42 ProGlnArgThrserserArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu 61
 DB 101 CCCAGAGAGAGCCCGTCGCGAGGTGCAGAGAGCTGACGCGCTTCGAGACTTC----- 154
 QY 62 AlaSerGlyLeuAsnAparThr-----AlaGlnProTrpThrAlaPro 75
 DB 155 CGGCGCGACTTCTGTGACAAACGACGCGGGGCTTCGAGAGACAGTGTACCGCGCGCG 214
 QY 76 LeuProLysGly-----LeuGluCysProValProAlaSerTyTrpAsnApplIePhe 92
 DB 215 CTGTGGAGATGACGCCACCGGACATGTCAGATTCCCTCCAGCTTCATATGATCATCAGC 274
 QY 93 IleSerArgGluIleHisAparHisValIGlyTrpValTyTrtyrGlnArgGluValIleVal 112

DB 275 CAGAGTGGCGTCTGGCGGCAATTTTTCGGCTGGGTGTGTACGAACGGAGGATCTCG 334
 QY 113 ProLYuGlyTrpSerGlnu-----ArgTyLeuValArgAlaGluSerAlaThr 129
 DB 335 CCGAGCCGATGAGCCAGAGCCTGCGCACAAAGTGTGCTGTGAGTGGCAGTCCCAT 394
 QY 130 HisHisGlyArgIleTyValAlaAsnAparGluLeuValAlaGlnHisValIGlyTyTrp 149
 DB 395 TCCTATGCCATGTGTGGTGAATGGGTGCAGACGCTTGAGCATAGAGGGGGCTTAC 454
 QY 150 ProPheGluAlaAspValThrGluLeuValAlaProGly-----GluTyPheArg 166
 DB 455 CCCTTGAGGCGGAGATGACAGAACTGTGCAGGTGGGGCCCTGCCGCTCCGCGTCCGA 514
 QY 167 LeuThrIleGlyValAlaAsnGlnLeuThrHisGlyThrIleProProGlyLysIle--- 185
 DB 515 ATCACTATTCGCGATGACAAACACATCACCCGCCACCACTGCGCACAGGAGCATCCA 574
 QY 186 -----ThrThrGlyAsnAlaThrGlyLysArgIleGlnThrTyTrpHisApar 201
 DB 575 TACCTGACTGACACTCCAAAGTATCCCAAGGTTACTTGTCCAAACATATTTTGAC 634
 QY 202 PheTyAsnTrpAlaGlyLeuAlaArgSerIleTrpLeuTySerValProGlnGlnHis 221
 DB 635 TTTTCACTAGCTGCTGAGTGCAGCGGTCTGTATCTGTACAGACACCCACCACTAC 694
 QY 222 IleGlnAparIleThrValAlaThrAparValAparGlyAparAsnGlyLeuLeuApar 241
 DB 695 ATCGATGATCATCACTCCATCCACCAAGCGTGGAGCAAGAGATGGGTGTGAATTCCAG 754
 QY 242 ValGluValAlaAsnGlnThrThrGlyGlnIleGlnIleSerValIleAparGluApar 261
 DB 755 ATCTGTGTCAAGGCGAGTAACTGTTCAAGTTGAAGTGTGCTTTGTGATGACGAAC 814
 QY 262 AlaIleValAlaAlaAsnAparGlyAlaGlnGlyThrValThrIleProSerValLysLeu 281
 DB 815 AAGTGTGGCGAATGGACCTGGAGCCAGGGCAACTTAAGGTGCAGGTGTCAAGCTTC 874
 QY 282 TrpGlnPro-----GlyAlaAlaTyLeuTyTrpGlnLeuGlnValAlaIle 296
 DB 875 TGGTGGCCGTACCTGATGACAGAACCCCTGCTATCTGTATTCATTTGAGTGGACGCT 934
 QY 297 ValGly-----SerSerGlyAparValAparThrTyAsnLeuAlaThrGlyValArg 314
 DB 935 ACTGCACAGACGATGAGGCGCTGTGTGATCTTACACATCTCCCTGTGGGAGATCCG 994
 QY 315 ThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysAparPheTyTrpThrGly 334
 DB 995 ACTGTGGCTGTACCAAGAGCCAGTTCATCATCAATGGGAAACCTTCTATTTCCACGG 1054
 QY 335 PheGlyLysHisGlnAparThrAlaValArgGlyLysGlyHisAparProAlaTyTrpMetVal 354
 DB 1055 GTCCAAACAACATGAGATCGGACATCCAGGGAAGGGCTTCGATCGGCGCTGTGGTG 1114
 QY 355 HisAparPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTyTrp 374
 DB 1115 AAGGACTTCAACTGCTGTGCTGGCTGTGGTGGCCAGCGTTCGATCCAGCCACTACCCC 1174
 QY 375 TYAlaGluGluValMetAparPheAlaAparArgAsnGlyIleValAlaIleAparGluThr 394
 DB 1175 TATGAGAGAGAGATGATGAGATGTGTGACCGGTATGGAGATTGTGTATTCATGATGAG 1234
 QY 395 ProAlaValGlyLeuAsnIleAlaLeuMetGlyValSerGluSerGlyAlaProGlnThr 414
 DB 1235 CCGCGCGTGGCGCTGCGCTG-----CCGAGTTC 1264
 QY 415 PheThrProAparAlaIleAsnAparLysThrGlnGluAlaHisGlyGlnAlaIleArgGlu 434
 DB 1265 TTC-----AACACGTTTCTGTGCATCACCACTGACGATGAGAA 1309
 QY 435 LeuIleAlaArgAparLysAsnHisAlaSerValAlaMetTrpSerIleAlaAsnGluPro 454

Db 1310 GTGGTGGTAGGAGCAAGAACACCCCGGTGATGATGCTGTGGCCAAACGACCT 1369
 Qy 455 AAlaserhsgluaspgrgluAlaarggluTyrPhegluProleuthrasnleuthraargin 474
 Db 1370 GCGGCCCACTAGATGTGCTGGCTACTCTT-GAA----- 1404
 Qy 475 LeuaspProthraargProleuthrPheAlaasnValGlyThralathrTyrGlnleuasp 494
 Db 1405 -----GCTCCGTA----- 1413
 Qy 495 ArgIleseraspheaspheaspValserCysGllleasnTyrPheGlyTyrPyrserGln 514
 Db 1414 -----GTGATGTGATCTGTGTGAAACGCTACTACTCTGTGATACGAC 1458
 Qy 515 ThrGlyAspHeugluGluAlaGluAlaAlaLeugluYsgluLeuHISGlyTyrGlnGlu 534
 Db 1459 TACGGGACACTGGAGTTGATTCAGCTGCAGCTGGCCACCCAGTTGAGAACTGGTATAG 1518
 Qy 535 LysPheHisArgProIleValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHIS 554
 Db 1519 AAGTATCAGAGCCCATTTATTCAGAGCGAGTATGGAGCAAAACGATTGGAGGGTTTAC 1578
 Qy 555 SerIleLeuGlyLeuProIlePyrserGluGluPheGlnValGlnMetLeuAspMetTyrHIS 574
 Db 1579 CAGGATCCACCTGTATGTCTACATAAGATACGAAAGTCTGCTAGACAGTATCCAT 1638
 Qy 575 ArgValPheAsp-----ArgIleGluSerMetAlaGlyGluHISValTyrAsnPheAla 592
 Db 1639 CTGGGCTGGATCAAAACGCGAGAAATACGTTGGAGAGTCATTTGGAAATTTTGCC 1698
 Qy 593 AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnGlyValPheThr 612
 Db 1699 GATTTCATGATCGAAGACATGCACGACGAGAGTCTGGGAAATAAAGGGGATCTTCAC 1758
 Qy 613 ArgAspArgLysProLysAlaAlaAlaHisSerLeuArgAlaArgTyrPheSerIle 631
 Db 1759 CGGACAGACACCAACCAAAAGTGCAGCGTCTTTTGGAGAGATCTGGAAGATT 1815

RESULT 3
 CD014092 1995 bp mRNA linear EST 21-OCT-2003
 LOCUS DEFINITION 90134867 Single gene library Homo sapiens cDNA, mRNA sequence.
 ACCESSION CD014092
 VERSION CD014092.1 GI:37777621
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1995)
 Jtn,P., Fu,G.K., Wilson,A.D., Yang,J., Chien,D., Hawkins,P.R.,
 Au-Young,J. and Stuve,L.L.
 PCR isolation and cloning of novel splice variant mRNAs from known
 drug target genes
 Genomics 83 (4), 566-571 (2004)
 CONTACT: Jtn, P.
 Incyte Corporation
 3160 Porter Drive, Palo Alto, CA 94304, USA
 Tel: 650 621 8639
 Fax: 650 621 8965
 Email: pjtn@incyte.com.

FEATURES

source
 1..1995
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="Single gene library"
 /note="Vector: pDrive Cloning Vector; RT-PCR was performed
 using gene-specific primers flanking the open-reading
 frame. PCR products were subcloned into pDrive Cloning
 Vector and sequenced completely using M13 forward and
 reverse primers. Sequencing gaps were closed by
 re-sequencing using primers flanking the gapped areas."

ORIGIN
 Alignment Scores:
 Pred. No.: 2,62e-120 Length: 1995
 Score: 1175.00 Matches: 251
 Percent Similarity: 55.92% Conservative: 89
 Best Local Similarity: 41.28% Mismatches: 181
 Query Match: 35.03% Gaps: 89
 DB: 6 Indels: 89
 Gaps: 12

US-10-757-093-4 (1-634) x CD014092 (1-1995)

Qy 42 ProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTyrPheAlaLeu 61
 Db 101 CCCAGAGAGAGCCCGTCGCGGAGTGCAAGAGCTGAGACGGCTTGAAGCTTC----- 154
 Qy 62 AlaSerGlyLeuAsnAspThr-----AlaGlnProTyrThraLaPro 75
 Db 155 CGGCGGACTTCTGTACAAACGAGCCGCGGCTTGAGAGCAAGTGTACCGCGCGCG 214
 Qy 76 LeuProLysGly-----LeuGluCysProValProAlaSerTyrAsnAspIlePhe 92
 Db 215 CTGTGGAGAGTCAAGCCCGCCACCGTGCAGATGCCAGTCCCTCAGCTTCATAGACATCAGC 274
 Qy 93 IleserArgGluIleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleVal 112
 Db 275 CAGAGTGGCGTCTGCGGCATTTGTGCGCTGGGTGTGTACGAAACGGAGAGTATCTCTG 334
 Qy 113 ProLysGlyTyrPyrserGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisHisGly 132
 Db 335 CCGAGAGCATGAGACCCAGAC----- 355
 Qy 133 ArgIleTyrValAsnAspArgLeuValAlaGluHisValGlyGlyTyrThrProPheGlu 152
 Db 356 -----CTGCCGACAAAGAGTGTG----- 373
 Qy 153 AlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsn 172
 Db 374 -----CTGAGGATTTGCAGTGC 391
 Qy 173 AsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlnAsnAlaThrGly 192
 Db 392 CATTC-CTATGCCATCTGTGATCTCCAAAG-GGTACTTT----- 428
 Qy 193 LysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIle 212
 Db 429 -----GTCCAGAACACATATTTTGACTTTTCAACTAGCTGTGAGCTGCAGCGTCTGTA 482
 Qy 213 TyrLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValValThrAspValAsp 232
 Db 483 CTTCGTATACAGACACCCACCTCATCATGATGATGATCACCTCCACCAAGCTGGAG 542
 Qy 233 GlyAspAsnGlyLeuLeuLeuLeuTyrGluValAlaGluValAlaAsnGlnThrThrGlyGlnIle 252
 Db 543 CAGACAGTGGGCTGGGATTAACAGATCTGTCCAGGCGAGTAACCTGTCAAGTTG 602
 Qy 253 GlnIleserValIleAspGluAspGlyAlaIleValAlaIleValAlaSerGlyAlaGlnGly 272
 Db 603 GAAGTGGCTTTTGGATGAGAAACAAAGTGTGGCAATGGAGCTGGAGCCAGGCGC 662
 Qy 273 ThrValThrIleProSerValLysLeuTyrGlnPro-----GlyAlaAla 287
 Db 663 CAACATTAAGGTGCAGAGTGTACGCTGTGTGGCGGTCTGATGCAGAACGCGCTGCC 722
 Qy 288 TyrLeuTyrGlnLeuGlnValAsnIleValGly-----SerSerGlyAspValValAsp 305
 Db 723 TATCTGTATTCATTTGAGAGTGCAGCTGCAGTGCAGACGTACCTGGGCGCTGTGTCTAC 782
 Qy 306 ThrTyrAsnLeuAlaThrGlyAlaArgThrValLysValAlaIleValSerGlnPheLeuIle 325
 Db 783 TTCATACACTCCCTGTGGGATCGCACTGTGGTGTACCAAGAGCACTTCTCTATC 842
 Qy 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGly 345

```

Db      843  AATGGGAAACCTTTCTATTCTTCCACGGGTGTCACAGAGATGAGGTGGACATCCGAGGG 902
Qy      346  LysGlyHHisAspProAlaTyrMetValHHisAspPheGlnMetLysTrpIleGlyAla 365
Db      903  AAGGGCTTCACACGGCCGCTGGTGGTGAAGAGACTTCAACTGCTTCCCTGGCTGGTGGCC 962
Qy      366  AsnSerPheArgThrSerHisTyrProTyrAlaGlnGluValMetAspPheAlaAspArg 385
Db      963  AACGCTTCCGTTCCAGCCACTACCCCTTATGCAAGAGATGATGCAAGATGTGTGACCC 1022
Qy      386  AsnGlyIleValValIleAspGlnThrProAlaValGlyLeuAsnIleAlaLeuMetGly 405
Db      1023  TATGGATGTGTGCTCATCATGATGATGTCGCCGGGTGGCTGGCGCTG----- 1070
Qy      406  ValSerGluSerGlyAlaProGlnThrPheThrProAspAlaIleAsnAspLysTrpGln 425
Db      1071  -----CCGACGTTCTTC-----AACCAAGTTCTCTG 1097
Qy      426  GLuAlaHleIleGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHleAlaSerVal 445
Db      1098  CATCACCACTGACGAGTGAAGAGTGTGCTGGAGCAAGACACCCCGCGTCT 1157
Qy      446  ValMetTrpSerIleAlaAsnGluProAlaSerHisGlnAspGlyAlaArgGluTyrPhe 465
Db      1158  GTGATGTGGTCTGGCGCAAGCGCTGCCCTCCACCTAGAACCTGTGCTGCTACTCTTG 1217
Qy      466  GluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsn 485
Db      1218  AAGATGTGTATCCCTCAACCAATCTTGTGACCCCTCCGCGCTGTGACCTTT----- 1271
Qy      486  ValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerGlyIle 505
Db      1272  GTACGCACTCTACTACTATGACAGACAGAAAGGGGCTCCGATGTGATGATGATCTGTTG 1331
Qy      506  AsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluGluAlaGluAlaLeu 525
Db      1332  AACAGCTACTACTCTGTATGACAGCTACGGGCACTGGAGTTGATGATTCAGCTGCAGCTG 1391
Qy      526  GluLysGluLeuHleGlyTyrGlnGluLysPheHisArgProIleValMetThrGlyTyr 545
Db      1392  GCCACCCAGCTTGAAGACTGTATTAAGATATCAGAACCCCTATTAATTCAGACGAGTAT 1451
Qy      546  GluAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluPhe 565
Db      1452  GAGGCGAAGAACGATGAGGGGTTTCACAGAGATCCACTGTGATGTTCACTGAAAGATAC 1511
Qy      566  GlnValGlnMetLeuAspMetTyrHisArgValPheAsp-----ArgIleGluSerMet 583
Db      1512  CAGAAAGCTGCTAGACAGCTACATCTGGCTGTGATCAAAAACGCAAAATACGTG 1571
Qy      584  AlaGlyGluHisValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
Db      1572  GTTGGAGACGCTCATTTGGCAATTTTGGCGATTCATCAACAGTACACGACGAGAGTG 1631
Qy      604  AspGlyAsnLysIleGlyValaPheThrArgAspArgLysProIleAlaAlaHisSer 623
Db      1632  CTGGGGAAATTAAGGGGATCTTCACTCGGACAGACCAACAAAGATGAGCGTTCCTT 1691
Qy      624  LeuArgAlaArgTrpThrSerIle 631
Db      1692  TTCCGAGAGAGATACCTGGAAGATT 1715

```

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

JOURNAL

99279253

MEDLINE

10349636

PUBMED

2

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

20499374

MEDLINE

11042159

PUBMED

3

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Komno, H., Akiyama, J., Nishi, K., Kikunai, T., Taahito, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, K., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-184-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

20530913

MEDLINE

11076861

PUBMED

4

AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the

PANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

JOURNAL

5

REFERENCE

The PANTOM Consortium and the RIKEN Genome Exploration Research

Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

JOURNAL

6

REFERENCE

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,

Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,

Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,

Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M.,

Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,

Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,

Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,

Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasumishi, A.,

Muramatsu, M. and Hayashizaki, Y.

Direct Submision

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp,

URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,

Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in Riken.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://phantom.gsc.riken.jp/

Location/Qualifiers

FEATURES

source

1..2274

/organism="Mus musculus"


```

Db      1625 CCGAACAAGAGGAGGATCTTCATCGCAGAGACAGCCCAAACTTCGGCTTATTTT 1684
Qy      624 uArgAlaArgTPrThSerIle 631
Db      1685 GCGAGAGAGATATCTGAGAGATT 1706

RESULT 5
CD014093      1853 bp      mRNA      linear      EST 21-OCT-2003
LOCUS      CD014093
DEFINITION      Single gene library Homo sapiens cDNA, mRNA sequence.
ACCESSION      CD014093
VERSION      CD014093.1 GI:37777622
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1853)
Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,
Au-Yang, J., and Stuve, L.L.
PCR isolation and cloning of novel splice variant mRNAs from known
drug target genes
Genomics 83 (4), 566-571 (2004)
CONTACT: Jin, P.
Incyte Corporation
3160 Porter Drive, Palo Alto, CA 94304, USA
Tel: 650 621 8639
Fax: 650 621 8665
Email: pj@incyte.com.
Location/Qualifiers
1. 1853
/organism="Homo sapiens"
/mol_type="mRNA"
/clone_xref="taxon:9606"
/clone_lib="Single gene library"
/notes="Vector: pDrive Cloning Vector; RT-PCR was performed
using gene-specific primers flanking the open-reading
frame. PCR products were subcloned into pDrive Cloning
Vector and sequenced completely using M13 forward and
reverse primers. Sequencing gaps were closed by
re-sequencing using primers flanking the gapped areas."

ORIGIN

Alignment Scores:
Pred. No.:      8,53e-106      Length:      1853
Score:      1045.50      Matches:      227
Percent Similarity:      50.90%      Conservative:      84
Best Local Similarity:      37.15%      Mismatches:      159
Query Match:      31.17%      Indels:      141
DB:      6      Gaps:      13

US-10-757-093-4 (1-634) x CD014093 (1-1853)
Qy      42 ProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu 61
Db      101 CCCGAGAGAGCCCGCGCGGAGTCCAGAGAGCTGAGCGGCTTCGAGCTTC----- 154
Qy      62 AlaSerGlyLeuAsnAspThr-----AlaGlnProTrpThraLPro 75
Db      155 CCGCGGCACTTCTGTGACAACGAGCGCGGCGCTTCGAGAGACAGTGTATCCGCGCGCGC 214
Qy      76 LeuProLysGly-----LeuGluCybProValProAlaSerTrpAsnAspIlePhe 92
Db      215 CTTGGGAGATCAGGCCCAACCGGTGACATGCCATGCTCCCTCCAGCTTCATATGACATCACC 274
Qy      93 IleSerArgGluIleHisAspHisValGlyTrpValTrpTrpGlnArgGluValIleVal 112
Db      275 CAGGACTGGCGCTGCGGCAATTTTGTGCGCTGGGTGTGTGTAAGACGAGAGGTGATCCG 334
Qy      113 ProLysGlyTrpSerGlnGlu-----ArgTrpLeuValArgAlaGluSerAlaThr 129
Db      335 CCGGAGCGATGAGCCAGAGACTCGCGCACAGAGTGTGTGCTGAGTGTGGCAGTGCC--- 391

```

```

Qy      130 HisHisGlyArgIleTrpValAsnAspArgLeuValAlaGluHisValGlyGlyTrpThr 149
Db      391 -----
Qy      150 ProPheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIle 169
Db      391 -----
Qy      170 GlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsn 189
Db      391 -----
Qy      190 AlaThrGlyLysArgIleGlnThrTrpGlnHisAspPheTrpAsnTrpAlaGlyLeuAla 209
Db      392 -----
Qy      210 ArgSerIleTrpLeuTrpSerValProGlnGlnHisIleGlnAspIleThrValAlaThr 229
Db      404 -----
Qy      230 AspValAspGlyAspAsnGlyLeuIleAsnTrpGluValAlaAlaAsnGlnThrThr 249
Db      410 -----
Qy      250 GlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGly 269
Db      452 TTCAGTGTGAAGTGGCGCTTTTGGATGCAAGAAACAAAGTGTGGCGAATGGAGCTGGG 511
Qy      270 AlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnPro----- 284
Db      512 ACCCAGGCCCACTTAAGTGTCCAGAGTGCACCTCTGTGGCCGTAACCTGATGACGAA 571
Qy      285 GlyAlaAlaTrpLeuTrpGlnLeuGlnValAsnIleValGly-----SerSerGlyAsp 302
Db      572 CGCCCTGCTATCTGTATTCATTTGAGAGGAGCTGACGTACGACAGACGTCACCTGGGCGCT 631
Qy      303 ValIleAspTrpThrTyraAsnLeuAlaThrGlyValArgTrpValLysValAlaGlySerGln 322
Db      632 GTGTCTGACTTTTACACACTCCCTGTGGGAGATCCGACCTGTGCTCCACCAAGACCG 691
Qy      323 PheLeuIleAsnGlyLysProPheTrpPheThrGlyPheGlyLysHisGluAspThrAla 342
Db      692 TTCCTCATCAATGGGAACCTTTCTATTTCCAGGGTGTCAACAGATGATGATGGCGAC 751
Qy      343 ValArgGlyLysGlyHisAspProAlaTrpMetValHisAspPheGlnLeuMetLysTrp 362
Db      752 ATCCGAGGGAAGGCGCTTCGACTGGCGCGTGTGTGAAGAGACTTCACCTGCTGCGCTGG 811
Qy      363 IleGlyAlaAsnSerPheArgThrSerHisTrpProTrpTrpAlaGluGluValMetAspPhe 382
Db      812 CTTGGGCCAAGCGCTTCCGTACACAGCCACTACCCCTATGCAAGAGAAATGATGACATG 871
Qy      383 AlaAspArgAsnGlyIleValValIleAspGluThrProAlaValAlaGlyLeuAsnIleAla 402
Db      872 TGTGACCGGTATGAGATGTGTGTATCATGATGATGATCCCGGCTGGCGCGCTG--- 928
Qy      403 LeuMetGlyValSerGluSerGlyAlaProGlnThrPheThrProAspAlaIleAsnAsp 422
Db      929 -----
Qy      423 LysThrGlnGluAlaHisLysGlnAlaIleAspGluLeuIleAlaArgAspLysAsnHis 442
Db      947 GTTTCCTGCATCACCACATGCGAGTGTGAGAAAGATGTGTGCTGAGGACAAAGAACAC 1006
Qy      443 AlaSerValAlaMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArg 462
Db      1007 CCGGCGGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1066
Qy      463 GluTrpPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThr 482
Db      1067 TACTACTTGAAGATGATGATGCTGCTACACCAATCTTGTGACCTCCCGGCTGTGACC 1126

```

Qy	483	h ³ ealaa ³ sn ³ l ³ g ³ yl ³ tn ³ l ³ a ³ th ³ r ³ g ³ l ³ ne ³ sp ³ ar ³ g ³ l ³ le ³ se ³ ar ³ le ³ u ³ p ³ he ³ a ³ sp ³ al	502
	1127	TTT-----GGAGCAACTTA ³ CTA ³ CTA ³ TA ³ GCACGA ³ CAAGGGGGCTCCGTA ³ TGCGAT ³ GC	1180
Qy	503	Se ³ r ³ Cys ³ l ³ lea ³ na ³ r ³ g ³ Y ³ p ³ he ³ g ³ l ³ Y ³ T ³ P ³ Y ³ se ³ r ³ h ³ n ³ h ³ rg ³ l ³ a ³ sp ³ leu ³ g ³ l ³ ua ³ gl ³ u	522
	1181	ATCTGTTGAACAGCTACTACTCTTGATCAGCACTACGGGCACTCGAGTTGATTTCAG	1240
Qy	523	Al ³ a ³ l ³ a ³ leu ³ g ³ l ³ u ³ Y ³ g ³ l ³ ue ³ h ³ i ³ g ³ l ³ Y ³ T ³ rg ³ l ³ ng ³ l ³ u ³ Y ³ p ³ he ³ h ³ sa ³ rg ³ Pro ³ l ³ e ³ Val ³ Me ³ c	542
	1241	CTGCAGCTGGCCACCACCTTTGAGAACTCGATTAAGAAGATGCAGAACCCATTATTACG	1300
Qy	543	Th ³ rg ³ l ³ u ³ Y ³ g ³ l ³ Y ³ l ³ a ³ sp ³ Th ³ leu ³ al ³ g ³ l ³ ue ³ h ³ i ³ se ³ r ³ l ³ e ³ u ³ g ³ l ³ Y ³ leu ³ Pro ³ T ³ ip ³ Se ³ r	562
	1301	AGCAGAT ³ GTGAGCAGAAACAT ³ TGCAGGGTT ³ CACAGCA ³ GTCCACCTCGATGTTTCACT	1360
Qy	563	Gl ³ u ³ g ³ l ³ u ³ p ³ he ³ g ³ l ³ ua ³ l ³ Me ³ l ³ e ³ u ³ sp ³ Me ³ c ³ Y ³ h ³ i ³ sa ³ rg ³ al ³ P ³ he ³ As ³ p-----Arg ³ L ³ e	580
	1361	GAAGGT ³ CCAGAAAAGCTCTGCTGAGCAGAT ³ CACTCGGGTCTGAT ³ CAAAAACGAGA	1420
Qy	581	Gl ³ u ³ se ³ r ³ Me ³ l ³ a ³ g ³ l ³ g ³ l ³ u ³ h ³ i ³ Val ³ T ³ ro ³ sp ³ he ³ l ³ a ³ sp ³ he ³ g ³ l ³ u ³ h ³ As ³ n ³ Leu ³ g ³ l ³ Y ³ l ³ e	600
	1421	AAATACGGGTGGAGACTCA ³ T ³ TGGAA ³ TTTTGCCAT ³ TTCAT ³ CA ³ GTCA ³ CTGAAC ³ CACTCACC	1480
Qy	601	l ³ le ³ ar ³ g ³ l ³ a ³ sp ³ g ³ l ³ Y ³ as ³ l ³ Y ³ le ³ g ³ l ³ Y ³ al ³ p ³ he ³ Th ³ ar ³ a ³ sp ³ ar ³ g ³ l ³ Y ³ Pro ³ Y ³ al ³ a ³ l ³ a	620
	1481	ACGAGAGT ³ GTGGGGAAT ³ AAAAAGGGA ³ TCTTCACTCGGCAGAGCA ³ CAAAAAGTCA	1540
Qy	621	Al ³ h ³ i ³ se ³ r ³ leu ³ al ³ g ³ l ³ ar ³ T ³ h ³ se ³ r ³ l ³ e	631
	1541	GGTTCCTTTTGGCAGAGAT ³ ACTCGAAGAT ³ T	1573

CR593823	1377 bp	mRNA	linear	HTC 21-JUL-2004
LOCUS	full-length cDNA clone CS0D001Y21 of B cells (Ramos cell line)			
DEFINITION	Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR593823			
VERSION	CR593823.1	GI:50474630		
KEYWORDS	HTC; CNSLT cDNA.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 1377)			
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished			
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue 2 (bases 1 to 1377)			
REFERENCE	Genoscope.			
AUTHORS	Direct Submission			
TITLE	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :			
JOURNAL	BP 191 91006. EVRY cedex - FRANCE (E-mail : seget@genoscope.cns.fr - Web : www.genoscope.cns.fr)			
COMMENT	1st strand cDNA was primed with a NotI-oligo(dt) primer. Five primers end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.			
FEATURES	Location/Qualifiers			
SOURCE	1..1377			

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODL001YM21"
/tissue_type="B cells (Ramos cell line)
/plasmid="pOWSPORT_6"

```

Alignment Scores:

Pred. No.:	8,286-95	Length:	1377
Score:	946.00	Matches:	186
Percent Similarity:	61.48%	Conservative:	63
Best Local Similarity:	45.93%	Mismatches:	130
Query Match:	28.21%	Indels:	26
DB:	3	Gaps:	6

QY 236 GlyLeuLleAsenThrgLnuValLnuValAlaAsnGlnThrTrgLygLnLleGlnLleSer 2555
 Db 1 GGGCTGGTGAATTAACCAAGATCTCTGTCAAGGGCGAGTAACCTGTTCAAGTTGGAAGTCGT 60
 QY 256 ValLleAspGluAspGlyValAlaLeValAlaIuAsalaserGlyAlaGlnGlyThrValThr 275
 Db 61 CTTTGGATGACGAAMAACAAGTCCGTGGGAAATGGAGTGGAGCCAGGGCCAAACTTAAG 120
 QY 276 IleProSerValIuLeuTrpGlnPro-----GlyAlaAlaIuTrLeuTr 290
 Db 121 GTGCCAGAGTGTACGCTCTGTGGCCGTAACTGATGACGAAGCCCTCGCTTAATCTGTAT 180
 QY 291 GlnLeuGlnValAsnLleValGly-----SerSerGlyAspValIleLaAspThrTrpAsn 308
 Db 181 TCATTTGAGGTGTGACGTGATCTGCACACAGCTGACTGGGGCTGTGTCTGACTTTCACA 240
 QY 309 LeuAlaThrGlyValAlaTrgThrValIuAsValAlaGlySerGlnPheLeuLleAsnGlyLys 328
 Db 241 CTCCTGTGTGGGATCCGCACGTGGCTGTCAACAAAGCCAGTTCCATCATATGGGAAA 300
 QY 329 ProPheTrpPheThrGlyPheGlyLysHisGluAspThrAlaValAlaArgGlyLysGlyHis 348
 Db 301 CCTTTCATTTCCACGGTGTCAACAAGCATGAGATGCGGACATCCGAGGGAGGGCTTC 360
 QY 349 AspProAlaTrpMetValHisAspPheGlnLeuMetTrpPileGlyValAlaAsnSerPhe 368
 Db 361 GACTGCGCGCTGCTGTGTGAAGACTTCAACCTGCTCTGCTGTGGTGTGACAGCTTTC 420
 QY 369 ArgThrSerHisTrpProTyrAlaGlnGluValMetAspPheAlaAspArgAsnGlyLe 388
 Db 421 CGTACCAAGCCACACACCTTATGCAGAGAAAGTATGACAGATGTGTGACCGCTATGGATT 480
 QY 389 ValValLleAspGluThrProAlaValGlyLeuAsnLleAlaLeuMetGlyValSerGlu 408
 Db 481 GTGGTCACTCATGATGATGCTCCGCGGTGGGCTGTGGCTGG----- 519
 QY 409 SerGlyAlaProGlnThrPheThrProAspAlaLleAsnAspLysThrGlnGluAlaHis 428
 Db 520 -----CCGACGATTTCTTC-----AACCAAGTTTCTCTGCATCACAC 555
 QY 429 LysGlnAlaLleArgGluLeuLleAlaArgAspLysAsnHisAlaSerValMetTrp 448
 Db 556 ATGCAGGTGATGGAAGAAGTGGTGGTGGGCAAGAACCCCGGGGTGTGATGTGG 615
 QY 449 SerLleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlyLysTrpPheGluProLeu 468
 Db 616 TCTGTGGCCAAAGAGCTGCTGCCCACTAGATCTGCTGCTACTACTTCGAAATGATGTG 675
 QY 469 ThrAsnLeuThrArgLnuLeuAspProThrAspProLleTrpPheAlaAsnValGlyThr 488
 Db 676 ATGCTGCACCAAAATCTTGAACCCCTCCGGCTGTGACCTT-----GTAGACAC 725
 QY 489 AlaThrTrpGlnLeuAspArgLleSerAspLeuPheAspValSerCysLleAsnAspTrp 508
 Db 730 TCTAACTATGACACAGACAAAGGGGCTCCGTATGTGTGATGTGATCTGTTGAACGCTAC 785
 QY 509 PheGlyTrpTrpSerGlnThrGlyAspLeuGlnGluAlaGluAlaLleuGlnLysGlu 528
 Db 790 TACTCTTGATATACAGACTACGGGCACTGTGAGTTGATTAACGTCGACCTGCGCACCCAG 845
 QY 529 LeuHisGlyTrpGlnGluLysPheHisAspProLleValMetThrGluTrpGlyAlaAsp 545
 Db 850 TTTCAGAACTGGTATGAAGATTAAGAAAGCCCATATTATTCAGAGCGAATATGAGACGAA 905

```

QY      549  ThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnValGln 568
DB      910  AGCATTTGACGGGTTTACACGAGATCCACCTGATGTTCATGACGAGATACGAGAAAGT 969
QY      569  MetLeuAspMetCysHisArgValPheAsp-----ArgIleGluSerMetAlaGlyGlu 586
DB      970  CTGCTGACGACAGATACATCTGGGTCTGATCAAAAAACGCAAAAAACGCTGGTGGAGAG 1029
QY      587  HisValTyrAspPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsn 606
DB      1030 CTCATTGGAAATTTCCCGATTCATGACTGAACACTGACCGACGAGAGTGCTGGGGAAT 1089
QY      607  LysLysGlyValPheThrArgAspArgLysProLysAlaAlaAlaHisSerLeuArgAla 626
DB      1090 AAAAAGGAGATCTTCTACCTCGGACAGACACAAAAGTCCACGTTCTTTGCCAGAG 1149
QY      627  ArgTyrThrSerIle 631
DB      1150 AGTACTGGAGAGATT 1164

RESULT 7
CD503076      1124 bp      mRNA      linear      EST 12-JUN-2003
LOCUS      CD460-C07.x1d-c SHGC-CDA Gasterosteus aculeatus cDNA clone
DEFINITION      CD460-C07 5', mRNA sequence.
ACCESSION      CD503076
VERSION      CD503076.1 GI:31432977
KEYWORDS      EST.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
ORGANISM      Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1124)
Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,
Schmutz, J. and Myers, R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
Contact: Kingsley, DM
HMMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@cmgm.stanford.edu
Plate: 60
High quality sequence stop: 782.
Location/Qualifiers
1..1124
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CD460-C07"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_1lb="SHGC-CDA"
/notes="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
(5' adaptor); Site 2: XhoI (3' linker primer). The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dt sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."

```

```

ORIGIN
Alignment Scores:
Pred. No.: 7,266-79      Length: 1124
Score: 803.00      Matches: 164
Percent Similarity: 60.74%      Conservative: 48
Best Local Similarity: 46.99%      Mismatches: 115
Query Match: 23.94%      Indels: 23
DB: 6      Gaps: 5

US-10-757-093-4 (1-634) x CD503076 (1-1124)
QY      284  ProGlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerGlyAspVal 303
DB      8  CGAACCAGGTTACTCTTACTCTTTGAGAGTTGCTTACTGCGCTGAGAGATCGCT 67
QY      304  Val--AspThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGln 322
DB      68  TCCACGACGTGTACCTCTACCAAGTCGACATCCGACGCTGACCTTACAGACCCAG 127
QY      323  PheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAla 342
DB      128  TTCCTCATCAACAAAAGCCCTTCTACTTCCACGGAGTAATTAACACAGAGCTGTGAT 187
QY      343  ValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrp 362
DB      188  ATTCCAGGCAAAAGCCCTGAGACTGCCCCCTCATGTGTAAGAGCTTTAATTGAAAGTGG 247
QY      363  IleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPhe 382
DB      248  TTGGGGGCGCAACTCGTTCCGACACAGCCACTCCCTTAAGCAGAGAGATCTCGCATAG 307
QY      383  AlaAspArgAsnGlyIleValValIleAspGluThrProAlaValGlyLeu---AsnIle 401
DB      308  TGTGACCGCCAGGACATCTGTGTGATGACGAGTGGCCGCGGTGGCATAAAGCAATT 367
QY      402  AlaLeuMetGlyValSerGluSerGlyAlaProGlnThrPheThrProAspAlaIleAsn 421
DB      368  CGCAGTTTGG----- 379
QY      422  AspLysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsn 441
DB      380  AACGCTCTTAACCATCAGTACCTGTCGTCATGACGACGACTGTGATCGTGGGCAAGAAC 439
QY      442  HisAlaSerValValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAla 461
DB      440  CATCCCTGTGTGTCATGTGTGATGAGCCCAATGACCCGCTGCACAGATGCTCTGTGCT 499
QY      462  ArgGluTyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIle 481
DB      500  GATTACTATTTCAAACCTTGATTAACATACCAAGATTGATTCACCCGCCCTGTC 559
QY      482  ThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAsp 501
DB      560  ACTTTT-----ATCACACAGAGTAATATGACAGAGATAAAGGGGCTCCCTACGTGAC 613
QY      502  ValSerCysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluGluAla 521
DB      614  GTCATCTGGCTAAACAGTTACTTCTCTGTCATGACCCCGGCGCACCCGAGATCATC 673
QY      522  GluAlaIleLeuGluLysGluLeuHisGlyTyrProlGlnGlyPheHisArgProIleVal 541
DB      674  CCCATCCAGCTCAACACTCAGTTTGAGAACTGTATCCGAAAGTACCAAGAAACCAATCATC 733
QY      542  MetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyr 561
DB      734  CAGAGGAGATTAAGGACGAGTGGGCGGCTTCAAGATGATCCACCCGATGTTT 793
QY      562  SerGluGluPheGlnValGlnMetLeuAspMetCysHisArgValPheAsp-----Arg 579
DB      794  ACTGAGAGTACAGAGAGTATGCTCGCAGAGCTTCCACACATGTTTCCACCAAGAAAGG 853
QY      580  IleGluSerMetAlaGlyLysHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599

```

Db 854 AGCAGTACGTCTACGCGCACTCTGCACTTGCAGACTTCATGACCCACAGG 913
 Qy 600 TlellarValaspGlyValPheThrArgAspArglySPolySala 619
 Db 914 ATATCGCGTGTGGGAACAAAGAGGTCTCTCGACAGGCA-AGGAGCCCAAGCAG 972
 Qy 620 AlaAlaHisSerLeuArgAlaArgTTP 628
 Db 973 CAGCATCTCTGAAGGAGAGTACTGG 999
 RESULT 8
 BX363460 1051 bp mRNA linear EST 08-APR-2004
 LOCUS BX363460 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
 DEFINITION Homo sapiens cDNA clone CS0DL001YM21 5-PRIME, mRNA sequence.
 ACCESSION BX363460
 VERSION BX363460.2 GI:46291531
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1051)
 Li,W.B., Gruber,C., Jesse,J. and Polyes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 5, 2003 this sequence version replaced gi:30372720.
 CONTACT: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 4169.r
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?c=CS0DL001YAG11QPlc=4169.r.
 Location/Qualifiers
 1..1051
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DL001YM21"
 /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
 /cell_line="RAMOS CELL LINE"
 /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES

source

ORIGIN
 Alignment Scores:
 Pred. No.: 2,72e-78 length: 1051
 Score: 797.50 Matches: 159
 Percent Similarity: 59.50% Conservative: 57
 Best Local Similarity: 43.80% Mismatches: 121
 Query Match: 23.78% Indels: 26
 Gaps: 6
 US-10-757-093-4 (1-634) x BX363460 (1-1051)

Qy 236 GlyLeuIleAsnTyrGluValAlaAsnGlnThrThrGlyGlnIleGlnIleSer 255
 Db 1 GGGCTGTGATTAACAGATCTCTGCAAGGCGAGTAACCTGTTCAAGTTGAAGGCGG 60
 Qy 256 ValIleAspGluAspGlyAlaIleValAlaIleSerGlyValGlnGlyThrValThr 275
 Db 61 CTTTGGATGCAAGAAACAAAGTCGTGGCAATGGAGCTGGGACCCAGGCGCAACTTAAG 120

Qy 276 IleProSerValIysLeuTTPPro-----GlyAlaAlaTyrIleuTyr 290
 Db 121 GGCCAGCGGTACAGCTCTGGTGGCGGTACGATGACAGAACGAGCCCTCATCTGAT 180
 Qy 291 GlnLeuGlnValAsnIleValGly-----SerSerGlyAspValValAspThrTyrAsn 308
 Db 181 TCATTGGAGTGCAGTGCAGTGCAGACAGACGTCGGGGCTGTGTCTCATCTTACACA 240
 Qy 309 LeuAlaThrGlyValArgThrValIysValAlaGlySerGlnPheLeuIleAsnGlyLys 328
 Db 241 CTCCCTGTGGGATCCGACCTGTGGCTGTGCACCAAGACCACTTCTCATATGGGAAA 300
 Qy 329 ProPheTyrPheThrGlyPheGlyLysHisGlyAspThrAlaValArgGlyLysGlyHis 348
 Db 301 CTTTCTATTTCACAGGTGTCAACAAGCATGAGATGCGGACATCGGAGGAAGGCTTC 360
 Qy 349 AspProAlaTyrMetValHisAspPheGlnLeuMetLysTTPIleGlyValAsnSerPhe 368
 Db 361 GACTGGCCGCTGCTGTGAGAGACTTCAACTGCTGTGCTGTGTGCTGTGCTGTGCT 420
 Qy 369 ArgThrSerHisTyrProTyrAlaGluValAlaMetAspPheAlaAspArgAsnGlyIle 388
 Db 421 CTTACAGCCACTACCTCTATGCAAGAGAGATGATGCAATGTGACCGCTATGGGATT 480
 Qy 389 ValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeuMetGlyValSerGlu 408
 Db 481 GTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 519
 Qy 409 SerGlyAlaProGlnThrPheThrProAlaAlaIleAsnAspLysThrGlnGluAlaHis 428
 Db 520 -----CCGACAGTCTTC-----AACACGGTTCCTTGATCACCAC 555
 Qy 429 LysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTyr 448
 Db 556 ATGCAAGTATGAGAGAGATGCTGCTGAGGACAGAACACCCCGGCTGTGATGTGG 615
 Qy 449 SerIleAlaAsnGluProAlaSerHisGlyAspGlyAlaArgGluTyrPheGluProLeu 468
 Db 616 TCTGTGGCCAGAGAGCTGCTGCCACCTGATGCTGCTGCTGCTGCTGCTGCTGCT 675
 Qy 469 ThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThr 488
 Db 676 ATGCTCACACCAAAATCCTTGGACCCCTCCGCTGTGACCTT-----GTGAGCAAC 729
 Qy 489 AlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyr 508
 Db 730 TCTAATATGCAAGCAGACAGAGGGGCTCGTATGTGATGTATCTGTTGAACAGCTAC 789
 Qy 509 PheGlyTTPTyrSerGlnThrGlyAspLeuGluGluAlaAlaLeuGluLysGlu 528
 Db 790 TACTCTGTGATACAGACATGAGGACCTGAGGATGATTACAGCTGAGCTGCCACCCAG 849
 Qy 529 LeuHisGlyTTPGlnGluLysPheHisArgProIleValMetThrGluTyrGlyAlaAsp 548
 Db 850 TTTGAGAACTGGTATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 909
 Qy 549 ThrLeu-AlaGlyLeuHisSerIleLeuGlyLeuProTTPSerGlnGluPheGluValGlu 568
 Db 910 ACGATTGGAGGATTTACACAGATCCACCTCTGATGTTCACTGTAAGACTAACAGAAAR 969
 Qy 568 nMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyLysHis-- 587
 Db 970 TTYTCTAGACAGTACCATCTGGTCTGTGATTAATAAAMGGAATAATATGTGTGGRRCT 1029
 Qy 588 -ValTTP 589
 Db 1030 CATTTGG 1036
 RESULT 9
 CL486845 921 bp DNA linear GSS 01-APR-2004
 LOCUS CL486845 Arabidopsis thaliana genomic clone
 DEFINITION SAIL_443_B06.v2 SAIL Collection

ACCESSION	TITLE
VERSION	JOURNAL
KEYWORDS	MEDLINE
SOURCE	PUBMED
ORGANISM	COMMENT
SAIL_443_B06.v2, genomic survey sequence.	
CL486845	
CY486845.1 GI:45965487	
GSS.	
Arabidopsis thaliana (thale cress)	
Arabidopsis thaliana	
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
rosoids; euroside II; Brassicales; Brassicaceae; Arabidopsis.	
1 (bases 1 to 921)	
Seesions,A., Burke,E., Presting,G., Aux,G., McEliver,J., Patton,D.,	
Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,	
Bullis,D., Snell,T., Miguel,T., Hutchison,D., Kimerly,B.,	
Mitxel,T., Katagiri,F., Glazebrook,K., Law,M. and Goff,S.A.	
A high-throughput Arabidopsis reverse genetics system	
Plant Cell 14 (12), 2985-2994 (2002)	
22356987	
12468722	
Contact: Seesions A	
Applied Trait Genetics	
Syngenta Biotechnology Inc.	
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA	
Email: allen.seesions@syngenta.com	
ABRC Stock Number CS820387; T-DNA left border flanking sequences of	
Syngenta Arabidopsis insertion library (SAIL) lines are available	
through the Arabidopsis Biological Resource Center (ABRC).	
Sequences represent a pool of amplified genomic regions and not	
single contiguous sequences.	
Class: TDNA tagged.	

FEATURES

```

Location/Qualifiers
1..921
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL_443_B06_v2"
/clone_lib="SAIL_Collection"
/note="T1-DNA left border sequences were isolated using a
modified TAIL-PCR strategy"

```

ORIGIN

Alignment Scores:	
Pred. No.:	2,51e-78
Score:	797.00
Percent Similarity:	76.45%
Best Local Similarity:	63.64%
Query Match:	23.76%
DB:	9
US-10-757-093-4 (1-634) x CL486845 (1-921)	
	Length: 921
	Matches: 154
	Conservative: 31
	Mismatches: 54
	Indels: 3
	Gaps: 2

Qy	388	llvalvalllleapsluThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyVal	406
Db	793	ATGTGTGATTAATGATTAACCTGCTGTGCGGCTTTAACTCTCTTTAGGCAATTGGCTTC	734
Qy	407	SerGluSerGlyAlaPro---GlnThr-PheThrProAspAlaIleAsnAspLysThrGly	425
Db	733	GAGCGGCGCAACAGCCGAAAGAACTGTACGCGAAGAGCGATCAACGGGGAACCTCA	674
Qy	425	ngluAlaHslybGlnAlaIleArgGlyLeuIleAlaArgAspLysAsnHsAlaSerVa	445
Db	673	GCAAGGCACTTACAGCGCATTTAAAGCTGTATACGGGTGACAAACCCACCAACCGT	614
Qy	445	lValMetTPSerIleAlaAsnGlyProAlaSerIleGlyAspGlyAlaArgGlyLysTrpH	465
Db	613	GGTGATGTGAGATTTGGCCACGACCCGATACCCGTCGCGAAGTGACACGGGAATTTT	554
Qy	465	eglupProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAs	485
Db	553	CGGCGCACATGGCGAAGCAACGGGTAAACTGCACCCGACGGCTCGATCACCTCGCAA	494
Qy	485	nValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIle	505

[illegible]

1
2
3
4
5
6

RESULT	10				
CD503098					
LOCUS			1138 bp	mRNA	linear
DEFINITION	CD503098				EST 12-JUN-2003
	CDNA60-D07.x1d-t SHGC-CDA Gasterosteus aculeatus CDNA clone				
ACCESSION	CD503098				
VERSION	CD503098.1				
KEYWORDS	EST,				
SOURCE	Gasterosteus aculeatus				
ORGANISM	Gasterosteus aculeatus (three spined stickleback)				

100

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1138)	Kingsley, D.M., Reichel, C., Balabhadra, S., Grimwood, J., Dickson, M., Schmutz, V. and Myers, R.M.	Expressed sequence tags from <i>Gasterosteus aculeatus</i>	Unpublished (2003)	Contact: Kingsley, DM

Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsleywcmgm.stanford.edu
Plate: 60
High quality sequence stop: 782.

FEATURES

```

source
1. .1138
   /organism="Gasterosteus aculeatus"
   /mol_type="mRNA"
   /strain="Salinas river, CA"
   /db_xref="taxon:69293"
   /clone="CDA60-D07"
   /sex="mixed male and female"
   /tissue_type="heads and internal organs combined"
   /dev_stage="adult"
   /clone_1fb="SHGC-CDA"
   /note="Vector: lambda ZAP Express/pBK-CMV; Site_1: EcoRI
(5' adaptor); Site_2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with

```

a 50 bp linker primer containing an oligo dt sequence preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector unidirectionally in the sense orientation with respect to the lacZ promoter of pBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual pBK-CMV phagemid clones for EST sequencing."

ORIGIN

Alignment Scores:

Pred. No.:	3.34e-75	Length:	1138
Score:	770.50	Matches:	166
Percent Similarity:	61.56%	Conservative:	47
Best Local Similarity:	47.98%	Mismatches:	111
Query Match:	22.97%	Indels:	25
DB:	6	Gaps:	5

US-10-757-093-4 (1-634) x CD503098 (1-1138)

```

QY 284 ProGlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspVal 303
DB 8 CGAACCAGGCTTACTCTTCTTGAAGGTTGCTTACTGCTGCTGAAGGATCTGCT 67
QY 304 Val--AapThrTyrAsnLeuAlaThrGlyValAlaGlySerGln 322
DB 68 TCCACGACGCTGACCTACAGCTGCGACGCGTCCGACGCTTACACGACCCAG 127
QY 323 PheLeuIleAsnGlyLeuProPheTyrPheThrGlyPheGlyLeuGlnAspThrAla 342
DB 128 TTCCTCATCAAAAGCCCTTCTTCTTCCAGGATTAATAAACAAGGAGCTCGAT 187
QY 343 ValArgGlyLeuGlyAspProAlaTyrMetValHisAspPheGlnMetLeuTyr 362
DB 188 ATTCGAGGCAAAAGGCTGCGACGCTGCTGAGGACCTTACTTATGAGAGTGG 247
QY 363 IleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGlnGluValMetAspPhe 382
DB 248 TTGGAGGCAAACTGTTCCGACGACGACCTTATGACGAGGAGATCTGACAGATG 307
QY 383 AlaAspArgAsnGlyIleValIleAspGlnThrProAlaValGlyLeu--AsnIle 401
DB 308 TGTGACCGGACATGCTGCTGATGACGAGGCGGCGCTGGGCAATAAAGACAT 367
QY 402 AlaLeuMetGlyValSerGlyValAlaProGlnThrPheThrProAspAlaIleAsn 421
DB 368 CGCAGTTTGA----- 379
QY 422 AspLysThrGlnGluAlaHisLysGlnAlaIleArgGlnLeuIleAlaArgAlaAsn 441
DB 380 AACCCCTCTTAACCCATCAGCTGCTGCTGACGACGCTGCTGCTGCTGCTGCT 439
QY 442 HisAlaSerValValMetTyrSerIleAlaAsnGlnProAlaSerHisGlnAspGlyAla 461
DB 440 CATCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 499
QY 462 ArgGlnTyrPheGlnProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIle 481
DB 500 GATTACTATTCAAAACCTTGATTAACATCAAAAGATTGATCAACCGGCGCGC 559
QY 482 ThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerSerPheLeuAsp 501
DB 560 ACTTTT-----ATCACAGACAGTACTATGCGAGGATTAAGGGGCTCTTACGTGAC 613
QY 502 ValSerCysIleAsnArgTyrPheGlyTyrSerGlnThrGlyAspLeuGlnGluAla 521
DB 614 GTGATCTGCGTAAACAGTACTCTCTGCTGACATGACCGGACACCGGAGGTATC 673
QY 522 GluAlaAlaLeuGlnLysGlnLeuHisGlyTyrGlnGlnLysPheHisArgProIleVal 541
DB 674 CCATCAGCTCAACACTGATTTGAGAACTGTACGGAAGTAACAGAAACCATCATC 733

```

```

QY 542 MetThrGlnTyrGlyAlaAspThrIleuAlaGlyLeuHisSerIleLeuGlyLeuProTyr 561
DB 734 CAGGCGCAATTAACAGAGGATGCGGCGGCTTCAAGTATCATCCACCGGATGTT 793
QY 562 SerGlnGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAsp-----Arg 579
DB 794 ACTGAGAGTACCAAGAACTTATGCTGACAGCTTACCAAGCTGTTCCAGCAAGAAAGA 853
QY 580 IleGlnSerMetAlaGlyGlnHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599
DB 854 AG-CAGTACGTCAATCGCGCAACTCATCTGGAAGTTGACAGATTATGACACAGAGG 912
QY 600 IleIleArgValAlaAspGlyLeuHisGlyValPheThrArgAspArgTyrGlyAla 619
DB 913 ATATCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 970
QY 620 AlaAlaHisSerLeuArg 625
DB 971 GCAGCATTCATCTTAAG 988

```

RESULT 11

BX745933/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 857
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TGAS066d19"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into PCS107 with Host: Escherichia coli XL1-blue.
Location/Qualifiers

ORIGIN

Alignment Scores:

Pred. No.:	1.9e-70	Length:	857
Score:	726.50 <td>Matches:</td> <td>144</td>	Matches:	144
Percent Similarity:	63.95%	Conservative:	44
Best Local Similarity:	48.98%	Mismatches:	85

```

Query Match:      21.66%      Indels:      21
DB:               5          Gaps:         5
US-10-757-093-4 (1-634) x BX745933 (1-857)

QY      283 GlnProGlyAlaAlaIleuTyGlnLeuGlnValAsnIleValGlySerSerGlyAsp 302
      849 GAGCCCTGGA-----TACTTGTACTCACTGAGATGATGATGATGAGCAAGATGCAAAAT 796
QY      303 -----ValValAspThrTyrAsnLeuAlaThrGlyValAlaArgThrValLysValAlaGly 320
      795 GGGTCAGTGAAGAAACATTAATTCCTTCCCTGAGAAATGAAGATGACGCTCTGGA 736
QY      321 SerGlnPheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHISGluAsp 340
      735 GACCAATTCCTCACTCAATGAGAAAACATTTACTTCCATGCTGTCACAAACATGAGAC 676
QY      341 ThrAlaValArgGlyLysGlyHISAspProAlaTyrMetValHISAspPheGlnLeuMet 360
      675 TATGATGTCAAGAGAAAGACTAGACTGCTCACTAATTTGTAAGATTTTAATCTGTTG 616
QY      361 LysTrpIleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGlnValMet 380
      615 AAGTGGCTTGCTTAATCTCTCCGACACGACATTAATCTTAATGCAAGAAATCATG 556
QY      381 AspPheAlaAspArgAsnGlyIleValIleAspGlnIleProAlaValGlyLeuAsn 400
      555 GACCTTGTGATTAATATGACATTTGCTGATGATGATGATGATGATGATGATGATGAT 496
QY      401 IleAlaLeuMetGlyValSerGluSerGlyAlaProGlnIleThrPheThrProAspAlaIle 420
      495 TAC-----CCCCAAGATTGTTGGAAACCAATCTTTA 466
QY      421 AsnAspLysThrGlnGlnAlaHISLysGlnAlaIleArgGlnLeuIleAlaArgLys 440
      465 AACAC-----CATTAATATGATGAGAGATTTGGCGCAGAGATGAA 421
QY      441 AsnHISAlaSerValIleMetTyrSerIleAlaAsnGluProAlaSerHISGluAspGly 460
      420 AACCGGCTTCCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361
QY      461 AlaArgGlnIleTyrPheGlnProLeuThrAsnLeuThrArgIleLeuAspProThrArgPro 480
      360 GCGGGGATATATTTAAACGGTATGTTAGTACCTTAAGCACTTGACCCACCCGACCA 301
QY      481 IleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPhe 500
      300 GTACACTAT-----GTTTCAATGCTTAATGATGATGATGATGATGATGATGATGATG 247
QY      501 AspValSerCysIleAsnArgTyrPheGlyTyrPheSerGlnIleThrArgIleAspLeuGlu 520
      246 GATGATATTTGTGAACAGTACTTTCTTTGTTGATGATGATGATGATGATGATGATGAT 187
QY      521 AlaGlnAlaLeuGlnLysGlnLeuHISGlyTyrPheGlnIleLysPheHISArgProIle 540
      186 ATCAACTCAACTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127
QY      541 ValMetThrGlyTyrGlyAlaAspThrLeuAlaGlyLeuHISerIleLeuGlyLeuPro 560
      126 ATACAGGTGAATATGAGACGATACAAATTCAGGGTTTCAAGTACGACCCCATTAATG 67
QY      561 TrpSerGlnIleGlnPheGlnValGlnMetLeuAspMetTyrHIS 574
      66 TTCACTGAAGATACAGAGAGTTGTGTGTAAGAAACTAATCAT 25

```

```

RESULT 12
LOCUS      CL679241
DEFINITION PRI0125C_C10_2 - PRI0125C_BR (811) Mixed stage fosmid library of P.
            pacificus var. California Pristionchus pacificus genomic, genomic
            survey sequence.
ACCESSION  CL679241
VERSION     CL679241.1 GI:50185841

```

```

KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
ORGANISM    Pristionchus pacificus
REFERENCE   1 (bases 1 to 811)
AUTHORS     Srinivasan,J., Otto,G.W., Kahlow,U., Geiseler,R. and Sommer,R.J.
TITLE       AppADB: an AcceDB database for the nematode satellite organism
            Pristionchus pacificus
JOURNAL     Nucleic Acids Res. 32 (1), D421-D422 (2004)
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@uebingen.mpg.de
            This library was generated at Caltech, Pasadena, USA and end
            sequenced at Vancouver, Canada.
            Seq primer: T7
            Class: fosmid ends.

FEATURES
    source
        1..811
            /organism="Pristionchus pacificus"
            /mol_type="genomic DNA"
            /strain="California"
            /db_xref="taxon:54126"
            /clone_lib="Mixed stage fosmid library of P. pacificus
            var. California"
            /note="Vector: pep1fos-5 Fosmid vector"

ORIGIN
Alignment Scores:
Pred. No.:      5,55e-69      Length:      811
Score:          712.50      Matches:      140
Percent Similarity: 67.51%      Conservative: 47
Best Local Similarity: 50.54%      Mismatches: 81
Query Match:    21.24%      Indels:      9
DB:             9          Gaps:         5
US-10-757-093-4 (1-634) x CL679241 (1-811)

QY      97 ILeHISAspHISValGlyTyrValTyrGlnArgGlnValIleValProLysGlyTyr 116
      1 ATTCGATTAATTAATGAGCAACGCTGCTATCAGCGGAAGTCTTATACGAAAGGTGG 60
QY      117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHISISeglyArgIleTyrVal 136
      61 GCAGGCGAGGATATCGTGCCTTTCGATGCGGTCACTCATTAACGCAAGTGTGGTTC 120
QY      137 AsnAsnArgLeuValAlaGlnHISValGlyTyrThrProPheGlnAlaAspValThr 156
      121 AATTAATCAGAAATGATGAGCATCAAGGCGGTATTAACGCAATTAAGCGATGTCACG 180
QY      157 GluLeuValAlaProGlnLysGlnLysPheArgLeuThrIleGlyValAsnAsnGlnLeuThr 176
      181 CCTATGTTATTCGCCGGAAGAACTGATGATGATGATGATGATGATGATGATGATGATG 240
QY      177 HISGlnThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196
      241 TGGCAGACTATCCCGCGGAATGATGATTAAC-----GACGAAACCGCAAGAAAGCAG 297
QY      197 ThrTyrGlnHISAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleThrLeuTyrSer 216
      298 TCTTACTTCATGATATTTCTTAACATGCGCGGATCATGCGGATTAATGCTTAACAC 357
QY      217 ValProGlnIleHISISeglnAspIleThrValIleThrAspValAspGly-----Asp 234
      358 ACCGCCAACAACCTGGGTGACATATCAACCGTGTATCAACCATCCGCAAGCCTTAAC 417
QY      235 AsnGlyLeuIleAsnTyrGlnValAlaGlnAlaAsnGlnThrThrGlyGlnIleGlnIle 254
      418 CAGCGCTGTTGACTGCGAGGTG---GTGGCAAT-----GATGATGTCACCGTT 465

```

Score:	705.50	Matches:	142
Percent Similarity:	61.44%	Conservative:	46
Best Local Similarity:	46.41%	Mismatches:	98
Query Match:	21.03%	Indels:	20
DB:	5	Gaps:	4
US-10-757-093-4 (1-634) x B0941196 (1-906)			
QY	331	TYRPHETHGLYPHEGLYUHEHISGLUHPHRIHALAVALARGGILYLVGLYHISASP	Pro 350
DB	2	TAATTCACGGGTGTCAACATCATAGATGGATGGACATCCAGGAGAGGGCTTGACTGG	61
QY	351	ALAATYMEVALHISAPRHEGLINDEUETLSTRIILEGILYAASNSERPHETGHR	370
DB	62	CCGGCTGCTGTGAAGGACTTCAACCTGCTGGCTGGCTGGTGCCAAACGCTTCCGTAAC	12
QY	371	SETHISTYRPROTYRALAGLUGILVALMECSAPRHEALAAAPRAGANGILYILEVALI	390
DB	122	AGCAGTACCCCTATATGCACAGAAATGTATGCAGATGTGTACCGCTATGGATGTGGTC	181
QY	391	ILASRGLUHPHPRALAVAGILEUENHILEALMEUGILVALSERGLUSERGLY	410
DB	182	ATGATGATGTGCCCGGCTGGGCTGGCGCTG-----	214
QY	411	ALAProGLINTHPHETHPROAPRALALEANSPRYSTHRINGILUALHISLEGILN	430
DB	215	---CCGCAAGTCTTC-----AACACGTTTCTGTGCATACCAATACAG	256
QY	431	ALAILAARGILUENILEALARGPRYAPRHNHISALASERVILALMETTPSERILE	450
DB	257	GTGATGAAAGATGTGTGTGGGACAAAGAACCCCGGCGTGTGATGTGTGTGTG	314
QY	451	ALASRGLUPROALASERHISGLUAPRGLYALARGILUTYRPHGLUPROLEUTHIRASN	470
DB	317	GCCAAGAGACTGGCGTCCACCTAGAACTGTGGCTGTACTTGAAGATGGATGGCT	376
QY	471	LEUTHIRARGILUENUSAPRTHIRARPROILETHRPHREALANVALGLYTHIRALTHR	490
DB	377	CACACCAAAATCCTTGGACCCCTCCGCGCTGTGACCTTT-----GTGAGCACTGTAC	430
QY	491	TYRGLINLEUSAPRGLISESERPHEUPHEAPRVALSERGYHILEANARGYRPHGLY	510
DB	431	TATGACGACACAGAGGGGCTCCGATGTGATGTGATGTGTGTGAACACTACTACT	490
QY	511	TRPTYRSEGLINTHRGILYAPRLEUGILUALAGILUALALEUGILUYGLILEUHNIS	530
DB	491	TGCTATACAGACATAGCGGACCTGGAGTTGATTACGTGACGTGGCCACCCAGTTTGAG	550
QY	531	GLYTPRINGILUYPRHNHISAPRPROILEVALMETHTRGILUTYRGLYALASRPTHLEU	550
DB	551	AACSTGTATPAAAGAAATATCAGAAACCCATTAATTCAGACCGAGTGTGGAGAAACATTT	610
QY	551	ALAGLILEUHNISERILEUENGLYEUPTPDSERGLUGILUPHEDINVALIGLMELEU	570
DB	611	GCAAGGTTTCAACAGGATCCACCTGTGATGTTCACSTGAAGAGTACACGAAAGCTGTGA	670
QY	571	ASPMETYRHHISARGVALPHEAPR-----ARGILEGUSERMELVALGLYHNIHISVAL	588
DB	671	GACAGACACCATCTGGCTGTGATCAAAAACGCAAAAATATATCGTGTGGAGACTGATT	730
QY	589	TRPASRPHREALASRPHGLINTHRASNLEUGILYILEARGVALAPRGLYASNLYS-LY	608
DB	731	TGGAAATTTTGCAGATTTCATGACTAACACGTACCGACGAAAGTCTGGGGGAAATATAA	790
QY	608	SGILVALPHETHRARGASPRGLYSPROVALHIALHISERILEUARGALARGTR	628
DB	791	GGGAGATCTTCACTCCGACAGACACAAACAAAAGTCCACGCTCTTTGGAGAGAGATA	850
QY	628	prhnSerIleapRlys 633	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia: Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 877)
Yi, Y., Desai, R., Olarte, M., Henthorn, P. and George A.L.
Expressed sequence tags from Canine heart
Unpublished (2003)
Other ESTs: CH3#042_503T3
Contact: George AL
Division of Genetic Medicine
Vanderbilt University
529 Light Hall, 2215 Garland Avenue, Nashville, TN 37232-0275, USA
Tel: 615 936 2660
Fax: 615 936 2661
Email: al.george@vanderbilt.edu
Insert length: 2032 Std Error: 0.00
Seq primer: T7: TAAACGACTCCTACATAGGAG
High quality sequence start: 39
High quality sequence stop: 859.
Location/Qualifiers

FEATURES

source
1..877
/organism="Canis familiaris"
/mol_type="mRNA"
/db_xref="taxon:9615"
/clone="CH3#042_503"
/issue_type="heart"
/cell_type="heart"
/dev_stage="mixed developmental stages (adult, 30 day - 40 day fetal)"
/clone_lib="Canine heart normalized cDNA library in pBluescript"
/note="Organ: heart; Vector: pBluescript; Site_1: 5' of vector NotI; Site_2: 3' of vector EcoRI; Tissue source: dog heart (adult, 30 day - 40 day fetal), right and left atria and ventricle. Dog breed - mixed (beagle, German shepherd, pointer, Irish setter). Library construction: oligo-dT primed"

ORIGIN

Alignment Scores:

Pred. No.: 1 54e-65 length: 877
Score: 683.00 Matches: 135
Percent Similarity: 62.54% Conservative: 47
Best Local Similarity: 46.39% Mismatches: 87
Query Match: 20.36% Indels: 22
DB: Gaps: 3

US-10-757-093-4 (1-634) x CF406520 (1-877)

QY 192 GILYLSARGILLEGINTHYRGINHLSAPPHEHYRANTYRALAGLYLEUALARGSER 211
DB 46 GGTACTTCGTCACAGACATACCTTGTGACTCTTCACTACCGGGCCCTGCATCGCCCT 105
QY 212 ILETREPLEUTYRSEVALPROGINGLNHISILEGLASPRILETHRVALLTHRAAPVAL 231
DB 106 GTGCTCTCTACACACACACTACTACTGACCTGACACATCACCGCTCACCGGCGGTG 165
QY 232 ASPGLYASPSANGLYLEUITLEASNTYRGILVAGLVALLAASNGINTHRTGILYGLN 251
DB 166 GACCAAGACACTGGGCTGTGGATACAGATTTCCTCCAGGGCAGGTGAACACTTCACG 225
QY 252 ILEGLNLESERVALLEASPGILUASPGILYALALEVALALALYSALASERGIYALAGLN 271
DB 226 CTGGAAGTATCTTCTTGATGAGGAAGCAAGGTCTGGCCCGGAGGACAGGAGCCAG 285
QY 272 GLYTHRVALLHRIEPROSERVALYSELUTDGLNPRO-----GLYALA 286
DB 286 GGCCTGTCAGAGGCCCAATGTCACCTGCTGGCCGTACTGATGATGACACACCCC 345
QY 287 ALATYREUTYRGINLEUGLNVALLASNILEVALGLY-----SERSEGLYASPVALL 304
DB 346 GCCTACTGTACTGCTGTGAGGTAGGCTAGCGGAGATGGCCGCTGGGCGCTGTCTCA 405
QY 305 ASPTHRYRANLEUALATHRGILYVALARGTHRVALLYSVALAGLYSERGLNPHLEU 324

DB 406 GACTTCTACTCTCCCGGTGGGATTCGACCGGTGGCCGTACAGAGCCGCACTTCCTC 465
QY 325 ILEANGLYLVSXPPOPHETYPHEHTRGLYPHEGLYLSHLSGLASPTHRALLAVALLARG 344
DB 466 ATCAAGGGGAACCTTTCTATTTCATGGGGTCAACAAACATGAGAGATGCCATATCCGA 525
QY 345 GILYSGILYHLSAPPSPROALATYRMEYALHLSAPPHEGLNLEUMETLYSTRPILLEGLY 364
DB 526 GGGAGGGGCTTGTGACTGGCCGCTGCTGGTGAAGAGACTTCAACTGTGCGCTGGCGGCG 585
QY 365 ALAASNSERPHETRYTHSERHSTRYTRPCTYRALAGLUGLVALLMETASPHENALAP 384
DB 586 GCCAATGCCCTTCGACACACGACCTACCCCTACGCGGAGAGTATGACGCTTCGCGAC 645
QY 385 ARGASNGLYLLEVALVALLASPGILHTRPROALVALLGLYLEUASNILEALALEUET 404
DB 646 CGCTATGGATGTGTGTCATGACAGAGAGCCCTGGTGTGGCATCATGCTGCTC----- 699
QY 405 GILVALSERGLUSERGLYALAPROGINTHRPHEHTRPROASPALALEASAPLYSTR 424
DB 700 -----CAGAGCTACGCAATGTGTCC 720
QY 425 GINGUALAHLSLYSGLNHALLEARGGILEUITLEARGASPLLYASPHLSLASER 444
DB 721 CTGCAAGCACCATCTGAGGTGATGGGAGCTGTGCGTGGATGAGATACCCATCT 780
QY 445 VALVALMETTRPSERILEALASNGILUPROALASERHLSGLASPGILYALARGILUTYR 464
DB 781 GTAGTCATGTGTCTGTACCAATGAGCCCACTTCTTCCTGAACCGCTGCTTACTAC 840
QY 465 PHEGLUPROLEUTHRASNULEUTHRARGGLNLEU 475
DB 841 TTCAAGACCGTATGCTCACCACCAAGGCCCTT 873

Search completed: March 22, 2005, 21:57:16
Job time : 4490 secs